BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Remarks

Claims 1-5, 11-13, 15-20 and 53-67 were pending and examined in the Office Action of March 8, 2004. Of those claims, 1-5, 11-13, 15-20, 53-55, 57-60, 62-67 stand rejected and 56, 61 and 66 are objected to. With this Amendment and Reply, claims 1, 2, 3, 13, 16, 18, 19 and 20 are amended, and claims 68-75 are newly added, to more particularly point out and distinctly claim the invention.

Applicants note with appreciation the withdrawal of the restriction between Groups I and II, and the finding that claims 56, 61 and 66 are only objected to.

Remarks on the rejections entered in the Office Action of March 8, 2004 follow.

Rejections under 35 U.S.C. 112, first paragraph

Claims 1-5, 11-13, 15-20, 54, 55, 57, 59, 60, 64, 65 and 67 stand rejected under 35 U.S.C. 112, first paragraph as not being described in the specification sufficiently to convey to the skilled artisan that the inventors had possession of the invention (written description rejection). It is asserted that the claimed functional variants are not sufficiently described in the specification. Applicant respectfully requests reconsideration and withdrawal of these rejections in light of the claim amendments, the enclosed Declaration of Bernard Hallet under 37 C.F.R. 1.132, and the following discussion.

Applicants first note that the claims as amended limit the functional variants within the scope of the invention to those that are altered from SEQ ID NO:3 only in the central crossover region. These functional variants are sufficiently described, e.g., at paragraph 61 on page 5 of the specification.

Applicants also assert that functional variants of TRT (SEQ ID NO:3) altered only in the central crossover region would be understood to be possessed by the inventors at the time of filing, since the literature at that time establishes that variants of a known recombination site for a tyrosine recombinase that vary only in the central crossover region would be expected to be functional. See enclosed Declaration of Bernard Hallet under 37 C.F.R. 1.132. As discussed therein, experimental data with tyrosine

recombinases, as well as the knowledge of the mechanism of tyrosine recombinase recombination, would have led a skilled artisan to conclude that the inventors had possession of such variants at the time of filing. For your convenience, we also enclose herewith as Exhibit B, References 2 and 6 cited in Dr. Hallet's Declaration, i.e., Van Duyne, 2001, Annu. Rev. Biophys. Biomol. Struct. 30:87-104, and Landy, 1989, Annu. Rev. Biochem. 58:913-949.

We also refer the PTO to Paragraphs 10 and 11 of Dr. Hallet's Declaration, as well as FIG. 1, attached thereto as Exhibit C, providing experimental results that confirm the expected outcome, that TRT variants in the central crossover region are substrates for the tyrosine recombinase TnpI.

Based on the above discussion and the Declaration of Dr. Hallet, it is clear that the skilled artisan would understand that the inventors had possession of the claimed invention, since the sequences of the finite number of variants of TRT that vary only in the central crossover region were clearly envisioned, and these variants would be expected to be functional as TnpI recombinase substrates. Applicants therefore respectfully request reconsideration and withdrawal of the rejections under 35 U.S.C. 112, first paragraph.

Rejections under 35 U.S.C. 102(b)

Claims 1-3, 11-13, 53, 55, 58 and 60 stand rejected under 35 U.S.C. 102(b) as being anticipated by Mahillon et al., 1988, NAR 16:11827. The PTO asserts that plasmid pGI2, disclosed therein, includes SEQ ID NO:3 and SEQ ID NO:2 but not more than 200 contiguous nucleotides of SEQ ID NO: 4. Applicants respectfully request reconsideration and withdrawal of this rejection, because the cited plasmid does contain more than 200 contiguous nucleotides of SEQ ID NO:4. Indeed, the entire SEQ ID NO:4 (249 nt) is in the pGI2 sequence at nt 498-746. Applicants also provide a printout of GenBank Accession X13481, which provides a more legible version of the pGI2 sequence from Mahillon et al., NAR 16:11827. All 249 nt of SEQ ID NO:4 is underlined therein. Since

pGI2 in Mahillon et al. does comprise the entire sequence of SEQ ID NO:4, that reference does not anticipate any of the claims. Accordingly, withdrawal of this rejection is respectfully requested.

Claims 1-3, 11-13, 53, 55, 58 and 60 also stand rejected under 35 U.S.C. 102(b) as being anticipated by Mahillon et al., 1988, EMBO J. 7:1515-1526. It is asserted that plasmids are described therein that comprise SEQ ID NO:3 and SEQ ID NO:2 but not more than 100 contiguous nucleotides of SEQ ID NO:4. Applicants respectfully request reconsideration and withdrawal of this rejection, because the only sequence of a plasmid provided in the cited reference is Tn4430, which comprises the entire SEQ ID NO:4, at nucleotides 1-249. For your convenience, the applicants also provide GenBank entry X07651, providing the Tn4430 sequence. The entire SEQ ID NO:4 sequence is underlined therein. There is also no indication that any of the other plasmids described comprise SEQ ID NO:3, but not the entire sequence of SEQ ID NO:4. Accordingly, withdrawal of this rejection is respectfully requested.

Claims 20, 63 and 65 stand rejected under 35 U.S.C. 102(b), as anticipated by, or 103(a), as obvious over Mahillon et al., 1988, EMBO J. 7:1515-1526. It is asserted that the claimed kits are anticipated or obvious because *E. coli* cells described therein that comprise the SEQ ID NO:2 or SEQ ID NO:3 containing vectors would be expected to also comprise both the Tnp1 protein. Applicants respectfully request reconsideration and withdrawal of these rejections based on the claim amendments and the following discussion.

The rejected claims as amended all require that the DNA molecule comprising one or more copies of TRT (SEQ ID NO:3), or functional variant, be in a separate container from the TnpI protein. This does not occur, or would not be expected to occur with any composition described in the cited reference because Mahillion et al., or any other reference published before the instant disclosure, does not suggest any function for the TRT sequence. Therefore, there would be no reason to have the isolated TRT sequence

or functional variant in the same kit, but in separate containers, as claimed. Accordingly, withdrawal of this rejection is respectfully requested.

Conclusion

Based on the claim amendments and the above discussion, applicants respectfully request withdrawal of all rejections and passage of the claims to allowance. If there are any minor matters preventing this result, applicants request that Examiner McGarry contact the undersigned attorney.

Applicants believe that no fee is required with this filing. However, if there are any unexpected fees required to maintain pendency of this application, the PTO is authorized to withdraw those fees from Deposit Account 01-1785.

Respectfully submitted

AMSTER, ROTHSTEIN & EBENSTEIN LLP
Attorneys for Applicant
90 Park Avenue
New York, NY 10016
(212) 336-8000

Dated: New York, New York
June 8, 2004

By: _____
Elie H. Gendloff
Registration No.: 44,704

X13481. Bacillus thuringi...[gi:3171732] Links

```
LOCUS       BTPGI2XX              9672 bp    DNA     linear   BCT 07-JUL-2002
DEFINITION  Bacillus thuringiensis plasmid pGI2 with transposon Tn4430.
ACCESSION   X13481
VERSION     X13481.1  GI:3171732
KEYWORDS    plasmid; plasmid pGI2; recombinase; resolvase; transposase;
            transposon; unidentified reading frame.
SOURCE      Bacillus thuringiensis
  ORGANISM  Bacillus thuringiensis
            Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus
            cereus group.
REFERENCE   1  (bases 1 to 6999)
  AUTHORS   Mahillon,J. and Seurinck,J.
  TITLE     Complete nucleotide sequence of pGI2, a Bacillus thuringiensis
            plasmid containing Tn4430
  JOURNAL   Nucleic Acids Res. 16 (24), 11827-11828 (1988)
  MEDLINE   89098342
   PUBMED   3211758
REFERENCE   2
  AUTHORS   Mahillon,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-NOV-1988) Mahillon J., Plant Genetics Systems, J
            Plateaustraat 22, B-9000 Gent, Belgium
  REMARK    revised by [3]
REFERENCE   3  (bases 1 to 9672)
  AUTHORS   Hoflack,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-1998) Hoflack L., Plant Genetics Systems, J
            Plateaustraat 22, B-9000 Gent, Belgium
COMMENT     On Jun 2, 1998 this sequence version replaced gi:40316.
FEATURES             Location/Qualifiers
     source          1..9672
                     /organism="Bacillus thuringiensis"
                     /mol_type="genomic DNA"
                     /strain="thuringiensis H1.1."
                     /db_xref="taxon:1428"
                     /plasmid="pGI2"
     repeat_region   493..4651
                     /transposon="Tn4430"
     repeat_region   493..497
                     /note="direct repeat 1"
     repeat_unit     498..535
                     /note="inverted repeat A"
     CDS             744..1598
                     /codon_start=1
```

/transl_table=11
/product="TnpI resolvase"
/protein_id="CAA31832.1"
/db_xref="GI:40317"
/db_xref="GOA:P10020"
/db_xref="Swiss-Prot:P10020"

/translation="MDVAKQFSSYLKQENKTENTVQGYTSGIRQYIKWFEGSYDRKLT

KLYRQNILEYISYLKNVKMLNAKSINHKISSLAKFNEFLIQKGSQQDQVILKTDMIKV

QTVYASPTQIVELDVKKFLQSVLEDNNKRNYAIATLLAYTGVRISEALSIKMNDFNLQ

TGECIIRSGKGGKQRIVLLNSKVLSAIKDYLIDRKTYSTAHESPYLFISKKREKLDRT

VVNRIFKSYSNVITPHQLRHFFCTNAIEKGFSIHEVANQAGHSNIHTTLLYTNPNQLQ
LKNKMELL"
CDS 1617..4580
/codon_start=1
/transl_table=11
/product="TnpA transposase"
/protein_id="CAA31833.1"
/db_xref="GI:40318"
/db_xref="GOA:P10021"
/db_xref="Swiss-Prot:P10021"

/translation="MGVKQLLSEAQRNELMDLSRLTEWDLVTFRTFSKHDLHLILKHR

RGYNRLGFALRLVLIRYPGWSLTEYKDIPQYVVAYVTSRLRIPPEEFLVYAKRGNTLW

EHLGEIRTEYGYQNFSSEYKETLLQFLVQQAMDNNNTLYLIEITISTLRKTKVILPAM

YVIEDIVWEAKQQADQKVYSILHDGLVQEQKDQLDALLLPTINGKSPLAWLKDVPAQP

SPESFLKVIDRLQFVQKIGLTIDTTKINTNRLRQLARLGSKYEPYAFRRFNEVKRYSM

LVSFLLEITQDLIDYAIEIHDRLMMNLQTKGKKEQDEIQQANGKKLNEKILQFITVCG

TLIEAKETGKDAFAALDEVMSWNEMVESVEEAKQLSRPLNYDYLDLLNTRYSYVRRYA

PTLLRSLHFRATKSGEPVLQALDTIHELNETGKRKVPHGAPLHFVSNRWQKHVYDDDG

NINRHYYELAALTELRNHIRSGDIFVSGSRHHKAFDDYLIPYDEWNEVSNIPNGLTAP

LKAEDYITDRINRLNEHLEWLSKNSEKLEGVDISQGKLHVERLDRGTPEEAKAFSKLL

HSMLPRIKLTDLLIEVASWTGFHDQFIHASTNQSPDQEEQNIVLATLMAMGTNIGLTK

MAEATPGISYRQMANASQWRMYDDAMVRAQSILVNFQKEQKLSSYWGDGTTSSSDGMR

LSIAVRSLHADSNPHYGTGKGGTIYRFVSDQLSAYHVKVITTNARDALHVLDGLLHHE

TDLKIEEHYTDTAGYTDQVFALTHLLGFRFAPRIRDLADTKLFSIPGGEEYENVQALL

KGKINVKLIKENYEDIRRLAYSVQTGKVSSALIMGKLGSYARQNKLATALGEMGRIEK

TLFTLDYISNKAVRRRVQKGLNKGEAINALARTIFFGQRGEFRERALQDQLQRASALN

IIINAISVWNTVYMEKAVEELKARGEFREDLMPYAWPLGWEHINFLGEYKFEGLHDTG
QMNLRPLRIKEPFYS"
repeat_unit 4609..4646
/note="inverted repeat A'"
repeat_region 4647..4651
/note="direct repeat 1"
CDS 4859..5473
/note="unnamed protein product; ORF 1"
/codon_start=1
/transl_table=11
/protein_id="CAA31836.1"
/db_xref="GI:3171733"
/db_xref="GOA:P10022"
/db_xref="Swiss-Prot:P10022"

/translation="MTNTIDFKHVEKNARIRDFENEKEKFKQDHNGINQEEVNQAMQV

LSKATGGKEIFIGTKRSPQSKVKFAQFIQDNWDYALENAFFTDEEMLFLLRIQRFLQF

KSNCIVNDIHSRNALPMSQKQIADRLKTDKSKISRIVNSLVQKGVIVKANGHKPEGVK
ARTYALFINPNIIYSGERDNVETTLKALFMNSKSLFKKFPIALF"
CDS 6037..6849
/note="unnamed protein product; ORF 2"
/codon_start=1
/transl_table=11
/protein_id="CAA31837.1"
/db_xref="GI:3171734"
/db_xref="GOA:P10023"
/db_xref="Swiss-Prot:P10023"
/translation="MKKLLISFIAILFFICGFNLKAFAAEEIIDYQSLYNQAIQEGVL

DQNSVSYNEWLKQNKEEFMPIYQDGLKQGVFLEPLSYNEWLKLNNYGQAPTGDIELFD

DVTPRGSWGGFTLKAGDIFITNATSSAGIVGHAAIANGDNYILHMPGAGQGNQQLSTS

NWMQKYTASGKWIKVYRLKDQTLARDVARYADRNFYSTTGSATKNVYLDYGIDTHLYQ
KNPTYCSKLVFQALYFGSGSRNVMQAVSGIVTPYGLIDTFTSAYRPSLVKTY"
CDS 6896..7240
/note="unnamed protein product; ORF 3"
/codon_start=1
/transl_table=11
/protein_id="CAA31834.1"
/db_xref="GI:40321"
/db_xref="GOA:P10024"
/db_xref="Swiss-Prot:P10024"
/translation="MMAIFISIVLVVSFVLFNRVQAKKSIYNYIARQGIQESQLKYID

FHKDFKMGGYWLAVYVEGENPDIYYEYSYQDKKVNFQAYFNSEKAIKKKMWGGSGLTE
IEMKKLKYPPLQ"
CDS 7829..9166
/codon_start=1
/transl_table=11
/product="put. GI2 site-specific recombinase"
/protein_id="CAA31835.1"
/db_xref="GI:40322"
/db_xref="GOA:P10025"
/db_xref="Swiss-Prot:P10025"

/translation="MNKFAIIHMQKFQISDVQGIQKHNQRQGKSKSNLDIDYSKSEQN

YDLLNQQKIRYESTIKQEISERVKRKPRANSVVLSEFVVTASPDYMHSLSLEEQKRYF

ESSLDFIQKRYGKQNTLYAMVHMDEATPHMHIGVMPITEDNRLSAKDMFTRKELISLQ

QDFPLEMREKGFDVDRGEGSEKKHLSPQAFKEKQDLEVEVEQLSNVKTHLKTKVVETH

NQLQQTTNYIEKQNETLQKIQQQFLSLDKKIKEKKQEFETFRNQIPDKPVSMSYLREE

TKTEVTTKLFGKPEITEKKTGNIVVTREQWRDMTEKVNAAVIVKKDYERLQKTDLVKE

NQSLREDNKYLEETIKGNNLALKHSYKQNRELEEVNKELHTEIGTLKAHIRDLQMNIK

VLYQQTKKVFKEQFKAFRGLIKNELDMKGVDNQFEREHTREIRSRQKGYDMER"
ORIGIN
1 caagactatg ttgtaaatgc tcatagcatt gtatgttgga tttcaattgt aagttccaaa
61 gtaaactttt gaatatttgt ttactttcaa ataattggtt ttatggtgtt ttcagaggat
121 tttagaaaaa tagagattcc cgtaaagaaa gtatacattt agaatgcgaa ttttgtacgg
181 atgatggtta aagatactca ataaaaattt cattttgtag aaaatcatca aaataaaacg

241 agcacttatg gcaaggcgct cgtttaaaaa acattctgtg tattagggtt gtgtttataa
301 gttatgaacc gatgaatatt tgttaaagta aggtttaaga tagtttattc attttaaagt
361 aaatgggaac ctgtacggca aaaaaagagg accttcgaac agtgctcttt ctctgagtct
421 gtagtgcatg atactttatg aagtttttt agaaataaaa tgaatgagct cacgactaaa
481 acgaacttgc tatacaaggg gtaccgccag catttcggaa aaaaaccacg ctaagaaaat
541 cagagttaaa aaatcagaaa atatatcatt attccaagac acatacagtg tctttttta
601 tacaaaaaat aatacaacac aatattaatt gtgttgtatt aggtgttata ataaatataa
661 atctaggggt ttaacgcaac acaatttatc gataaataaa tacttttaga cgcaacacaa
721 tttatagacg cggaggaaat cacatggatg ttgcaaaaca gttttcttct tatcttaaac
781 aagagaataa aaccgagaac actgttcagg gatacacatc aggtattaga cagtacataa
841 aatggtttga aggttcctat gacagaaaat taacaaaatt gtaccgacaa aatatcttag
901 agtacattag ttatttaaag aatgtcaaaa tgttgaacgc caagtccatt aaccacaaga
961 ttagtagcct tgctaaattt aatgaatttc taatacagaa aggaagtcaa caagatcaag
1021 taattttaaa aacagacatg ataaaggttc aaactgtcta tgcttctcca acccaaattg
1081 ttgaattaga tgtaaaaaag tttttacaaa gtgtgttaga ggataataac aaacgtaatt
1141 atgcaattgc cactctccta gcatatacag gagtacgtat ttcagaggca ttatctatca
1201 aaatgaatga cttcaattta cagactgggg aatgtattat tcgaagtgga aaaggaggta
1261 aacaacgaat tgtattacta aatagtaagg tacttagtgc tatcaaagat tatctcatcg
1321 atcgaaaaac atacagtaca gcacatgaat ctccgtatct ttttattagt aaaaagcgag
1381 aaaagctcga ccgtacggtc gtcaatcgta tctttaaatc atacagcaat gttattactc
1441 cacaccaatt acgacacttc ttctgtacga atgcaattga aaaaggattt agcattcatg
1501 aagttgcaaa tcaagctggg cactctaaca tccatacgac actactttac acaaatccaa
1561 accaactgca gctaaaaaat aaaatggagc tcttataaag gagtaaaggg gaaacaatgg
1621 gagttaaaca attactgtca gaggcacaac gaaatgagtt aatggattta tcacgtttaa
1681 cagaatggga tttggtaaca tttcgtactt tttctaaaca tgacttacat ttaatattaa
1741 aacaccgtag aggatacaat cgtttagggt tcgctcttcg gcttgtatta attcgttatc
1801 ctggctggtc cttaacggaa tataaggaca ttccacaata tgtagtggct tacgttacga
1861 gtcggttacg gattccacct gaagaatttt tggtatatgc caaacgtggg aatactttat
1921 gggagcatct tggggaaatt cgaacagaat atggatatca gaatttttct tctgaatata
1981 aagaaacttt attacaattt ttagtgcaac aagcaatgga taacaataat accctctatt
2041 taatcgaaat aacaatatct acattgcgaa aaacgaaagt cattcttcca gctatgtatg
2101 tgattgaaga tatcgtctgg gaagcaaagc aacaagcaga tcagaaagtg tatagtatac
2161 tacatgacgg tttggtccaa gaacaaaaag atcaattaga tgcgttactt ttaccaacta
2221 taaatggtaa gtctccatta gcatggctaa aagatgtacc ggcccagcct tcaccagaat
2281 catttttaaa agtaatagac cgattacaat ttgttcaaaa aattggtctt actattgata
2341 ccacaaaaat taatacaaat cgtcttcggc agctcgctag attaggatca aaatatgagc
2401 catatgcttt tagacgattc aatgaagtaa aaagatatag tatgttagtc tcatttttat
2461 tggagattac acaagacctt atagattacg caattgaaat tcatgatcgt cttatgatga
2521 atcttcaaac aaaagggaaa aaagaacaag acgaaataca gcaagcaaat gggaaaaaat
2581 taaacgagaa gattttacaa tttattactg tatgcgggac actaattgag gcaaaagaaa
2641 cgggtaaaga tgcttttgca gccctagatg aagtaatgtc atggaatgag atggtcgaat
2701 ctgttgaaga agcaaaacaa ctatcacggc ctttaaatta tgactatttg gatttgttaa
2761 atactcgtta ttcctatgtt agacgatatg cacctacttt attacgcagt cttcatttta
2821 gagcgacgaa atctggagaa cccgttttac aagcacttga tacgattcat gaattgaatg
2881 aaacaggtaa acgaaaagtg cctcatggcg ctcccttaca ctttgtatct aaccgttggc
2941 aaaaacatgt atatgatgat gacggaaaca ttaatcggca ctattatgaa ttggcagctt

3001 taaccgagtt aagaaatcat attcgttcag gagatatttt tgtatcaggc agtagacatc
3061 ataaggcatt tgatgattat cttataccgt atgatgaatg gaacgaagtt tcaaacattc
3121 caaatggcct tacagcccct ttaaaagctg aagattatat aactgaccga ataaaccgat
3181 taaatgaaca tttagaatgg ttatcaaaaa atagcgagaa attagaagga gtagatatta
3241 gtcaaggaaa actacatgtt gagcgattgg atagaggaac tccagaagaa gcaaaagcat
3301 ttagtaaact tcttcatagt atgttaccca gaataaaact gacagattta ctaatagaag
3361 tggccagttg gactggattt catgatcaat ttatccacgc ctctaccaat caatctcctg
3421 atcaagagga acagaatatt gtattagcta ctttgatggc aatggggaca aatataggtc
3481 ttacaaaaat ggcggaagct acccctggaa tctcctatag acaaatggca aatgcttcac
3541 aatggagaat gtatgatgat gcaatggttc gcgctcaatc tattctcgtg aatttccaaa
3601 aggaacaaaa gttatcatcc tattggggag atgggacaac ttcttcttct gatggtatgc
3661 gtttatccat tgctgtacgt tctctacatg cagattccaa tccacactat ggaacgggaa
3721 aaggtggaac catttatcga tttgtaagtg atcagctatc tgcctatcat gtgaaagtca
3781 ttactaccaa tgcacgagat gctcttcatg tattagatgg gcttcttcat catgaaacag
3841 atttaaaaat tgaagaacat tacacagata cagctggata tacagatcaa gttttcgctt
3901 taacacactt gttgggattt cgatttgctc ctcgtattcg tgatttggca gatactaagc
3961 ttttttctat acctggtgga gaagaatatg aaaatgttca agcactttta aaaggaaaaa
4021 ttaatgtaaa gttaataaaa gaaaactatg aagatattag aaggttggca tactcggttc
4081 aaacaggaaa agtatctagc gctcttatca tggggaagct cggatcatat gcaagacaaa
4141 ataaactcgc aacagcactt ggagaaatgg gacgaatcga gaaaaccctc tttacgctgg
4201 attacatatc taataaagcc gtaaggagac gggttcaaaa aggtttaaat aaaggagaag
4261 caatcaatgc attagctaga actatatttt ttggacaacg tggagaattt agagaacgtg
4321 ctctccaaga ccagttacaa agagctagtg cactaaacat aattattaac gctataagtg
4381 tgtggaacac tgtatatatg gaaaaagccg tagaagaatt aaaagcaaga ggagaattta
4441 gagaagattt aatgccatat gcgtggccgt taggatggga acatatcaat tttcttggag
4501 aatacaaatt tgaaggatta catgacactg ggcaaatgaa tttacgtcct ttacgtataa
4561 aagagccgtt ttattcttaa tataacggct ctttttatag aaaaaatcct tagcgtggtt
4621 tttttccgaa atgctggcgg taccctaca ataacatata aaaagtcaga caagatacaa
4681 caaaaaaagc cattcatgct ataattgagt tgtcgaatct acaactatag gagaagggct
4741 ttttttactg ttatacaatt aattagagtg ttcagactca aattaattta tctaacttca
4801 taataacaaa ataaaagtct cttcgctagt tgtaaatact agaaacgagg agaaaagcat
4861 gacaaatacc attgatttta aacatgttga gaaaaatgca agaatacgtg attttgagaa
4921 tgaaaaagaa aagtttaaac aagatcataa cggaattaat caagaagaag taaatcaagc
4981 aatgcaagta ttaagtaaag ctactggcgg aaaagaaata tttataggta ctaaacgttc
5041 cccgcaatca aaagtaaagt ttgcccaatt tattcaagat aattgggatt atgctttaga
5101 aaatgcattt tttactgacg aagaaatgct ctttttactt cgtattcaaa ggtttttaca
5161 atttaaaagt aattgtattg taaatgatat tcattctaga aatgcacttc ctatgagtca
5221 aaaacaaata gctgatagat taaaaactga taaatctaaa attagtcgca ttgtaaatag
5281 tctggtacaa aaaggtgtta ttgtaaaagc aaatggacat aaaccagaag gggttaaagc
5341 tcgtacatac gcattattta ttaatcctaa tattatttat agcggtgaac gtgataatgt
5401 tgaaacaact ttaaaagctt tatttatgaa ttcgaaatca ctatttaaaa aatttccaat
5461 agccctttt taagggctat ttttttaatt ttaatagttg tctaatttga caactaaagt
5521 tgtctaatct gacaactatt aaaaccacta taaacccttg atataataag ctttttaact
5581 agtttagtag gtattttact ctcactctct atataagggc taaaagttcc gcaaggaaca
5641 ggtcgtattt ttgaggaaaa agcactcaaa aatctccacc tttttcacta cgtttcatcc
5701 ttgccaaact tttttcggct ggtttttagc tttcgccaag ccgaacagac caaaaacggg

5761 gttaagttgc tcgacttggg ctagcaattg taccagccga accgccaatt cttccgttca
5821 aagcatgatt ctttttccat atgtatcaac gctatgtggg gtttagcccc cacacgacta
5881 gccagcactc acacccaaaa acaagggaga attacattat ttagctaata aatatttgta
5941 tgtttttgt cacaagtatt tacattttga atttgaaaat cgtgttacct ttataaaagg
6001 ttaccgaaat ataataatta tttagggaga caatttatga aaaaattatt aattagtttt
6061 attgctattt tatttttat ttgtggtttt aacttgaaag cgtttgctgc tgaggaaatt
6121 attgattatc aatctcttta taatcaagca attcaagaag gtgttttaga tcaaaatagt
6181 gtatcttata atgaatggtt aaaacaaaat aaagaagaat ttatgcctat atatcaagat
6241 ggtttaaaac aaggtgtttt tctagaaccg ttatcttata atgaatggtt aaaattaaat
6301 aattatggac aagcaccaac gggagatatt gaacttttg atgacgtgac tccaagagga
6361 agttggggtg gatttacttt aaaagctgga gatatattta taacaaacgc aactagttca
6421 gcaggtatag taggacatgc agctattgct aatggtgata attatatttt acatatgcct
6481 ggtgctggcc aaggaaatca acaattatct acatctaatt ggatgcaaaa atatactgcg
6541 tcaggaaaat ggattaaggt ttatcgatta aaagatcaaa cacttgctag agatgtagct
6601 agatatgctg atagaaattt ttattccact actggaagtg caactaaaaa tgtttattta
6661 gattatggta tcgatacaca tctgtatcaa aaaaatccaa cttattgttc taaattagta
6721 tttcaagcac tctattttgg ttctggttca agaaatgtta tgcaagcggt tagtggaatc
6781 gtaactcctt atggtttaat tgatactttt acttcagcat atagaccgtc attagtaaaa
6841 acatattaat aaataattta ggaggattga atttgaaaaa aaaaagatgt taataatgat
6901 ggctatcttt atatcaatag ttttagtggt tagttttgtt ttatttaata gagtacaggc
6961 gaaaaaatct atatataatt acattgctag acaaggaata caagaaagtc agttgaaata
7021 tattgatttt cataaagact ttaaaatggg cggatattgg cttgctgttt atgtggaagg
7081 agaaaatcct gatatttatt atgaatattc ttatcaagat aaaaaagtta attttcaagc
7141 ttactttaat tctgaaaaag cgataaagaa aaaaatgtgg ggtggaagtg gacttactga
7201 aatagagatg aaaaaattaa agtatccacc tttgcaataa taaacagtat aaaaaagaac
7261 ttaaatacag gataaaacaa agactaatta gtactaaata gataagtact aattagtctt
7321 tttatttttt aattaattta gaattaagag tttgataaag gatttaatga atgattgaaa
7381 taactcgcaa aacgactatt gggagaagaa aaataagcgg taagtgtcaa tggtaagcgg
7441 actggaatgg ttcagcggtt tatgctgatc catgaaagga gcattaagga atggacacgt
7501 gcaaccgatt ttccatcctc ttggtagttc ctcaaaacga accgttttgg ggacgccaac
7561 cggcgaggaa gcccctcaag gattgagggg ttggggagtt ggatcgatgg ggtcaagggg
7621 aagagttccc cttgtggggt gtggggcaaa gcccctggtc tttatcacgc tgtggcgtga
7681 tttggcgaag ccaacgaggt ctcgcagagc ccacaccaaa tgtaatttgg tataatgggc
7741 tataccaaaa ctttggattc gctctctcga tatgttatga taagctcaaa agtagtcctc
7801 ttacggaaga caattgtgag ggaaaaacat gaacaaattt gcaatcattc acatgcagaa
7861 atttcaaatt tctgatgtgc aagggattca aaaacataat caaagacaag gaaaaagtaa
7921 atcaaattta gacatcgatt attcgaaaag tgaacagaat tatgatttgt tgaatcaaca
7981 aaaaattaga tatgaaagca caattaaaca agagatcagt gagcgagtaa aacgaaagcc
8041 acgggcaaat tctgttgtgc tttcagaatt tgtagtaaca gcttcgcctg attacatgca
8101 ttcattaagt ttagaagaac agaaacgtta ttttgagagt tcactagatt ttattcaaaa
8161 gcgttatggc aaacaaaata cgttgtatgc gatggttcat atggatgaag cgacaccaca
8221 tatgcatatt ggtgtgatgc caattacaga agacaatcgt ttgtccgcaa aagatatgtt
8281 tacgagaaaa gaattgatat ctttgcaaca ggacttcccg ttggaaatgc gggaaaaagg
8341 ttttgatgta gatcggggag aaggttctga gaaaaaacat ttatcgcctc aagcctttaa
8401 ggaaaaacaa gatttagaag tagaagtgga acagttatcc aatgtaaaaa cgcatttaaa
8461 gacaaaagta gtggaaacgc ataatcaatt acaacaaacc acaaactaca tcgaaaaaca

8521 gaatgaaacg ttacaaaaaa ttcagcaaca atttttgagt ttagataaaa aaataaaaga
8581 aaagaaacag gaatttgaaa cgtttcgaaa tcaaattccg gataagccag tgtcaatgtc
8641 ttatttgcga gaagaaacaa aaacagaagt aacaacgaaa ctgtttggaa agccagaaat
8701 tactgagaag aaaacaggaa acattgtggt tacacgtgag caatggagag atatgacaga
8761 aaaagtgaat gcagccgtta ttgttaaaaa agattatgaa cgtttacaga agacggatct
8821 tgtaaaagag aatcaaagtt tacgagaaga taataaatat ttagaagaaa caataaaggg
8881 gaataatctt gctttaaaac attcttataa acaaaatcgg gaattagaag aagtaaataa
8941 agaattgcat actgaaatag gcactttaaa ggctcatata agagatttac agatgaatat
9001 aaaggtttta tatcaacaga cgaaaaaagt ctttaaagag caatttaagg catttagagg
9061 acttattaaa aacgaactgg atatgaaggg tgtagataat caatttgaac gtgaacatac
9121 aagagaaata agaagtagac agaaaggata tgacatggaa cgttagaaac gttttgtgat
9181 atactataat tattcatcac taactaattt atgggtttca tccatttaaa aatctcttta
9241 tatgtgtgca tataaagaga ttttttgcgt ttatgtaatt aaaattatgt taagattgat
9301 aagactacat ttgaaaatgt gggttcaagt tgctgtttga ctgaattgca caaaaaacca
9361 attctattac acgagaattg gttttttgc gtataacgaa tagagatata tctaggcgtt
9421 taaggagttg tatcttatat aatttagtct aacattaata aaaagaaaat tcaaaaatat
9481 tcactaggtt tgttgtttat aaggaattat aaaacatgaa attttgcata ttttctaatt
9541 aattcccgga aataggaatt aatgatttat aataaataga tgaaataccc tttcagtttc
9601 cgatgaactg tactttagtc tgttgttagt acgaagtctt ttaaaagact tcgtactttt
9661 tttattgtcg ac

LOCUS BTTN4430 4149 bp DNA linear BCT 12-SEP-1993
DEFINITION Bacillus thuringiensis transposon Tn4430.
ACCESSION X07651
VERSION X07651.1 GI:40347
KEYWORDS plasmid; resolvase; tnpA gene; tnpI gene; transposase; transposon.
SOURCE Bacillus thuringiensis
ORGANISM Bacillus thuringiensis
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus; Bacillus cereus group.
REFERENCE 1 (bases 1 to 4149)
AUTHORS Mahillon,J. and Lereclus,D.
TITLE Structural and functional analysis of Tn4430: identification of an integrase-like protein involved in the co-integrate-resolution process
JOURNAL EMBO J. 7 (5), 1515-1526 (1988)
MEDLINE 88312602
PUBMED 2842151
COMMENT Data kindly reviewed (03-APR-1989) by Lereclus D.
FEATURES Location/Qualifiers
source 1..4149
/organism="Bacillus thuringiensis"
/mol_type="genomic DNA"
/strain="H1.1"
/db_xref="taxon:1428"
/map="plasmid pGI2"
/clone="pGI200"
/clone_lib="pBR322"
repeat_unit 1..38
/note="inverted repeat A"
CDS 247..1101
/note="unnamed protein product; put. resolvase Tnp I (AA 1 - 284)"
/codon_start=1
/transl_table=11
/protein_id="CAA30491.1"
/db_xref="GI:40348"
/db_xref="GOA:P10020"
/db_xref="Swiss-Prot:P10020"

/translation="MDVAKQFSSYLKQENKTENTVQGYTSGIRQYIKWFEGSYDRKLT

KLYRQNILEYISYLKNVKMLNAKSINHKISSLAKFNEFLIQKGSQQDQVILKTDMIKV

QTVYASPTQIVELDVKKFLQSVLEDNNKRNYAIATLLAYTGVRISEALSIKMNDFNLQ

TGECIIRSGKGGKQRIVLLNSKVLSAIKDYLIDRKTYSTAHESPYLFISKKREKLDRT

VVNRIFKSYSNVITPHQLRHFFCTNAIEKGFSIHEVANQAGHSNIHTTLLYTNPNQLQ
LKNKMELL"
CDS 1120..4083
/note="unnamed protein product; transposase Tnp A (AA 1 -
987)"
/codon_start=1
/transl_table=11
/protein_id="CAA30492.1"
/db_xref="GI:40349"
/db_xref="GOA:P10021"
/db_xref="Swiss-Prot:P10021"

/translation="MGVKQLLSEAQRNELMDLSRLTEWDLVTFRTFSKHDLHLILKHR

RGYNRLGFALRLVLIRYPGWSLTEYKDIPQYVVAYVTSRLRIPPEEFLVYAKRGNTLW

EHLGEIRTEYGYQNFSSEYKETLLQFLVQQAMDNNNTLYLIEITISTLRKTKVILPAM

YVIEDIVWEAKQQADQKVYSILHDGLVQEQKDQLDALLLPTINGKSPLAWLKDVPAQP

SPESFLKVIDRLQFVQKIGLTIDTTKINTNRLRQLARLGSKYEPYAFRRFNEVKRYSM

LVSFLLEITQDLIDYAIEIHDRLMMNLQTKGKKEQDEIQQANGKKLNEKILQFITVCG

TLIEAKETGKDAFAALDEVMSWNEMVESVEEAKQLSRPLNYDYLDLLNTRYSYVRRYA

PTLLRSLHFRATKSGEPVLQALDTIHELNETGKRKVPHGAPLHFVSNRWQKHVYDDDG

NINRHYYELAALTELRNHIRSGDIFVSGSRHHKAFDDYLIPYDEWNEVSNIPNGLTAP

LKAEDYITDRINRLNEHLEWLSKNSEKLEGVDISQGKLHVERLDRGTPEEAKAFSKLL

HSMLPRIKLTDLLIEVASWTGFHDQFIHASTNQSPDQEEQNIVLATLMAMGTNIGLTK

MAEATPGISYRQMANASQWRMYDDAMVRAQSILVNFQKEQKLSSYWGDGTTSSSDGMR

LSIAVRSLHADSNPHYGTGKGGTIYRFVSDQLSAYHVKVITTNARDALHVLDGLLHHE

TDLKIEEHYTDTAGYTDQVFALTHLLGFRFAPRIRDLADTKLFSIPGGEEYENVQALL

KGKINVKLIKENYEDIRRLAYSVQTGKVSSALIMGKLGSYARQNKLATALGEMGRIEK

TLFTLDYISNKAVRRRVQKGLNKGEAINALARTIFFGQRGEFRERALQDQLQRASALN

IIINAISVWNTVYMEKAVEELKARGEFREDLMPYAWPLGWEHINFLGEYKFEGLHDTG

QMNLRPLRIKEPFYS"
repeat_unit 4112..4149
/note="inverted repeat A"'
ORIGIN
1 ggggtaccgc cagcatttcg gaaaaaaacc acgctaagaa aatcagagtt aaaaaatcag
61 aaaatatatc attattccaa gacacataca gtgtcttttt ttatacaaaa aataatacaa
121 cacaatatta attgtgttgt attaggtgtt ataataaata taaatctagg ggtttaacgc
181 aacacaattt atcgataaat aaatactttt agacgcaaca caatttatag acgcggagga
241 aatcacatgg atgttgcaaa acagttttct tcttatctta aacaagagaa taaaaccgag
301 aacactgttc agggatacac atcaggtatt agacagtaca taaaatggtt tgaaggttcc
361 tatgacagaa aattaacaaa attgtaccga caaaatatct tagagtacat tagttattta
421 aagaatgtca aaatgttgaa cgccaagtcc attaaccaca agattagtag ccttgctaaa
481 tttaatgaat ttctaataca gaaaggaagt caacaagatc aagtaatttt aaaaacagac
541 atgataaagg ttcaaactgt ctatgcttct ccaacccaaa ttgttgaatt agatgtaaaa
601 aagtttttac aaagtgtgtt agaggataat aacaaacgta attatgcaat tgccactctc
661 ctagcatata caggagtacg tatttcagag gcattatcta tcaaaatgaa tgacttcaat
721 ttacagactg gggaatgtat tattcgaagt ggaaaaggag gtaaacaacg aattgtatta
781 ctaaatagta aggtacttag tgctatcaaa gattatctca tcgatcgaaa aacatacagt
841 acagcacatg aatctccgta tctttttatt agtaaaaagc gagaaaagct cgaccgtacg
901 gtcgtcaatc gtatctttaa atcatacagc aatgttatta ctccacacca attacgacac
961 ttcttctgta cgaatgcaat tgaaaaagga tttagcattc atgaagttgc aaatcaagct
1021 gggcactcta acatccatac gacactactt tacacaaatc caaaccaact gcagctaaaa
1081 aataaaatgg agctcttata aaggagtaaa ggggaaacaa tgggagttaa acaattactg
1141 tcagaggcac aacgaaatga gttaatggat ttatcacgtt taacagaatg ggatttggta
1201 acatttcgta ctttttctaa acatgactta catttaatat taaaacaccg tagaggatac
1261 aatcgtttag ggttcgctct tcggcttgta ttaattcgtt atcctggctg gtccttaacg
1321 gaatataagg acattccaca atatgtagtg gcttacgtta cgagtcggtt acggattcca
1381 cctgaagaat ttttggtata tgccaaacgt gggaatactt tatgggagca tcttggggaa
1441 attcgaacag aatatggata tcagaatttt tcttctgaat ataaagaaac tttattacaa
1501 tttttagtgc aacaagcaat ggataacaat aataccctct atttaatcga aataacaata
1561 tctacattgc gaaaaacgaa agtcattctt ccagctatgt atgtgattga agatatcgtc
1621 tgggaagcaa agcaacaagc agatcagaaa gtgtatagta tactacatga cggtttggtc
1681 caagaacaaa aagatcaatt agatgcgtta cttttaccaa ctataaatgg taagtctcca
1741 ttagcatggc taaaagatgt accggcccag ccttcaccag aatcattttt aaaagtaata
1801 gaccgattac aatttgttca aaaaattggt cttactattg ataccacaaa aattaataca
1861 aatcgtcttc ggcagctcgc tagattagga tcaaaatatg agccatatgc ttttagacga
1921 ttcaatgaag taaaaagata tagtatgtta gtctcatttt tattggagat tacacaagac
1981 cttatagatt acgcaattga aattcatgat cgtcttatga tgaatcttca aacaaaaggg
2041 aaaaaagaac aagacgaaat acagcaagca aatgggaaaa aattaaacga gaagatttta
2101 caatttatta ctgtatgcgg gacactaatt gaggcaaaag aaacgggtaa agatgctttt
2161 gcagccctag atgaagtaat gtcatggaat gagatggtcg aatctgttga agaagcaaaa
2221 caactatcac ggcctttaaa ttatgactat ttggatttgt taaatactcg ttattcctat
2281 gttagacgat atgcacctac tttattacgc agtcttcatt ttagagcgac gaaatctgga
2341 gaacccgttt tacaagcact tgatacgatt catgaattga atgaaacagg taaacgaaaa
2401 gtgcctcatg gcgctccctt acactttgta tctaaccgtt ggcaaaaaca tgtatatgat
2461 gatgacggaa acattaatcg gcactattat gaattggcag ctttaaccga gttaagaaat

2521 catattcgtt caggagatat ttttgtatca ggcagtagac atcataaggc atttgatgat
2581 tatcttatac cgtatgatga atggaacgaa gtttcaaaca ttccaaatgg ccttacagcc
2641 cctttaaaag ctgaagatta tataactgac cgaataaacc gattaaatga acatttagaa
2701 tggttatcaa aaaatagcga gaaattagaa ggagtagata ttagtcaagg aaaactacat
2761 gttgagcgat tggatagagg aactccagaa gaagcaaaag catttagtaa acttcttcat
2821 agtatgttac ccagaataaa actgacagat ttactaatag aagtggccag ttggactgga
2881 tttcatgatc aatttatcca cgcctctacc aatcaatctc ctgatcaaga ggaacagaat
2941 attgtattag ctactttgat ggcaatgggg acaaatatag gtcttacaaa aatggcggaa
3001 gctacccctg gaatctccta tagacaaatg gcaaatgctt cacaatggag aatgtatgat
3061 gatgcaatgg ttcgcgctca atctattctc gtgaatttcc aaaaggaaca aaagttatca
3121 tcctattggg gagatgggac aacttcttct tctgatggta tgcgtttatc cattgctgta
3181 cgttctctac atgcagattc caatccacac tatggaacgg gaaaaggtgg aaccatttat
3241 cgatttgtaa gtgatcagct atctgcctat catgtgaaag tcattactac caatgcacga
3301 gatgctcttc atgtattaga tgggcttctt catcatgaaa cagatttaaa aattgaagaa
3361 cattacacag atacagctgg atatacagat caagttttcg ctttaacaca cttgttggga
3421 tttcgatttg ctcctcgtat tcgtgatttg gcagatacta agctttttc tatacctggt
3481 ggagaagaat atgaaaatgt tcaagcactt ttaaaaggaa aaattaatgt aaagttaata
3541 aaagaaaact atgaagatat tagaaggttg gcatactcgg ttcaaacagg aaaagtatct
3601 agcgctctta tcatggggaa gctcggatca tatgcaagac aaaataaact cgcaacagca
3661 cttggagaaa tgggacgaat cgagaaaacc ctctttacgc tggattacat atctaataaa
3721 gccgtaagga gacgggttca aaaaggttta aataaaggag aagcaatcaa tgcattagct
3781 agaactatat tttttggaca acgtggagaa tttagagaac gtgctctcca agaccagtta
3841 caaagagcta gtgcactaaa cataattatt aacgctataa gtgtgtggaa cactgtatat
3901 atggaaaaag ccgtagaaga attaaaagca agaggagaat ttagagaaga tttaatgcca
3961 tatgcgtggc cgttaggatg ggaacatatc aattttcttg gagaatacaa atttgaagga
4021 ttacatgaca ctgggcaaat gaatttacgt cctttacgta taaaagagcc gttttattct
4081 taatataacg gctcttttta tagaaaaaat ccttagcgtg gttttttcc gaaatgctgg
4141 cggtacccc

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Application No. : 09/895,435 Confirmation No. 8975

Applicant : A. Francis Stewart, Youming Zhang and Bernard Hallet

Filed : June 30, 2001

Art Unit : 1635

Examiner : Sean McGarry

Docket No. : 29248/17

Customer No. : 1912

DECLARATION OF BERNARD HALLET UNDER 37 C.F.R. 1.132

Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Sir:

I, Bernard Hallet, declare that:

1. I received a Ph.D. degree from The Université Catholique de Louvain in 1993.

2. I presently hold the position of Research and Professor Associate at the Unité de Génétique of the Institut des Sciences de la Vie at the Université Catholique de Louvain. I have been a professor at the Université Catholique de Louvain since 2001.

3. My complete academic background, publications and professional experience are set forth in my Curriculum vitae, a copy of which is attached hereto as Exhibit A.

4. I have authored or co-authored more than 18 scientific articles in the area of genetic engineering.

5. I am a co-inventor of the invention claimed in U.S. Patent Application No. 09/895,435. I have reviewed the contents of this patent application, the Examiner's Office Action dated March 8, 2004, and the claims as amended in response to that Office Action. I have reached certain conclusions regarding the amended claims based upon that review and my knowledge of the field of art and my experience.

6. Amended claims 1, 13, 16 and 20 provide for the use of a "DNA molecule comprising one or more copies of TRT (SEQ ID NO:3) or a functional variant thereof that is altered from SEQ ID NO:3 only in the central crossover region". A molecular biologist with knowledge of the tyrosine recombinase literature and U.S. Patent Application No. 09/895,435 would understand that functional variants altered from SEQ ID NO:3 only in the central crossover region would be substrates for TnpI-mediated recombination. All changes in the central region (between the TRT inverted repeats) should remain functional (not only those that make the recombination site symmetrical). This understanding is based on the following discussion.

7. In current mechanistic models for recombination mediated by tyrosine recombinases, DNA strands are exchanged between the recombination partners by swapping and reannealing a few nucleotides from the central, or "crossover" region (6 to 8 bp) that separates the top and bottom strand cleavage positions within the recombination core sites (1, for a review, see Ref 2). There is no need for specific contacts between the recombinase protein and the central sequence during strand exchange.

8. As a consequence of this mechanism, tyrosine recombinases can function with recombination site variants containing altered crossover regions, provided that the identity between the partners is preserved. This was demonstrated in early studies on model recombinases such lambda Int and Cre (e.g., Ref. 3-5; reviewed in Ref. 6). However, changes in the central region was shown affect the overall efficiency of

recombination by influencing the cleavage reaction, the order of strand exchange, or the isomerization of the Holliday junction intermediate (e.g., Ref. 7-11). The ability to work on different pairs of recombination sites with a single recombinase is exploited by the Xer system to process dimeric forms of the bacterial chromosome and a variety of naturally occurring plasmids (For a review see Ref. 12). Early experiments also demonstrated that for simple recombination systems such as Flp/FRT or Cre/loxP, the orientation of the recombination site is wholly determined by the central region, recombination between symmetrical sites giving rise to both inversion and deletion events (e.g., Ref. 13-16).

9. References cited in Paragraphs 7 and 8:

1. Nunes-Duby SE, Azaro M, and Landy A (1995) Curr. Biol. 5:139-148

2. Van Duyne GD (2001) Annu. Rev. Biophys. Biomol. Struct. 30:87-104.

3. Weisberg RA, Enquist LW, Foeler C and Landy A (1983) J. Mol. Biol. 170:319-342.

4. DeMassy B, Studier F, Dorgai L, Appelbaum E and Weisberg RA (1984) Cold Spring Harbor Symp. Quant. Biol. 49:715-726.

5. Bauer CE, Gardner JF, Gumport RI (1985) J. Mol. Biol. 181:187-197.

6. Landy A. (1989) Annu. Rev. Biochem. 58:913-949.

7. Azaro MA and Landy A (1997) EMBO J. 16:3744-3755.

8. Hallet B, Arciszewka LK, and Sherratt DJ (1999) Mol. Cell 4:949-959.

9. Arciszewska LK, Baker RA, Hallet B and Sherratt DJ (2000) J. Mol. Biol. 299:391-403.

10. Lee L, and Sadowski, PD (2001) J. Biol. Chem. 276:31092-31098.

11. Lee L, and Sadowski, PD (2003) J. Mol. Biol. 326:397-412.

12. Sherratt DJ, Arciszewska LK, Blakely G, Colloms S, Grant K, Leslie N and McCullogh R (1995). Philos. Trans. R. Soc. Lond. B. Biol. Sci. 347:37-42.

13. Andrews BJ, McLeod M, Broach JR and Sadowski PD (1986) Mol. Cell Biol. 6:2482-2489.

14. Hoess RH, Wierzbicki A, and Ambreski K (1986) Nucl. Acids Res. 14:2287-2300.

15. Senecoff JF, and Cox MM (1986) J. Biol. Chem. 261:7380-7386.

16. Meyer-Leon L. Huang LC, Umlauf SW, Cox MM, and Inman RB (1988) Mol. Cell Biol. 8:3784-3796.

10. We have also confirmed that the claimed functional variants of SEQ ID NO:3 are substrates for TnpI-mediated recombination. To demonstrate this, functional variants of the TRT minimal core recombination site (SEQ ID NO:3) were constructed which included sequences in which the central 4-bp sequence between the inverted repeats was modified. In particular, TnpI-mediated recombination of a DNA substrate carrying two TRT sites with a bidirectional symmetrical central region was demonstrated, giving rise to both inversion and deletion products (See FIG. 1). In addition, changes in the central region were found to influence the overall efficiency of recombination, with the TRT variant containing the 5'-ATAT-3' central region being more efficient than the TRT variant with the 5'-TATA-3' central region (FIG. 1).

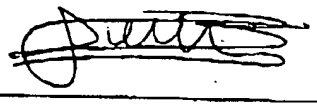
11. The legend of FIG. 1 is as follows: TnpI-mediated recombination at symmetrical core sites in vitro. (A) Nucleotide sequence of the wild-type (1.2) and symmetrical (1.1 and 2.2) core recombination sites. Changes in the 4-bp central region are shown in bold. Positions of TnpI cleavages are indicated. (B) In vitro recombination activity was examined on supercoiled plasmids carrying two copies of the different versions of the TRT sequence. After phenol extraction and ethanol precipitation, reactions were cleaved with SmaI and run on a 0.8% agarose gel. Bands are identified as follows : S1 and S2, substrate fragments; P1 and P2, deletion products; I1 and I2, inversion products; Int1 and Int2, product bands arising from intermolecular recombination reactions between inversely oriented substrates that cannot be seen with the wild type substrate; HJ, Holliday junction; Lin, linear form of the substrate.

12. Based on the above review of the literature, a molecular biologist who was provided with the present patent application would have understood at the time of filing that the claimed functional variants would serve as substrates for TnpI-mediated

recombination, and that the inventors would have contemplated such variants. The experimental results provided herewith confirm the expectation that such functional variants would be substrates for TnpI-mediated recombination.

13. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Signed,



Bernard Hallet

03/06/2004

Date

BERNARD HALLET

né le 28 octobre 1966 à Bruxelles

Belge, marié, père de trois enfants

ADRESSES

Professionelle

Université Catholique de Louvain
Institut des Sciences de la Vie (ISV)
Unité de Génétique (GENE)
Place Croix du Sud, 5 bte 6
B-1348 Louvain La Neuve
Belgique
Tel. (0032) 10 47 91 60
Fax. (0032) 10 47 31 09
Email Hallet@gene.ucl.ac.be

Privée

Rue Humerée, 50
B-5140 Sombreffe
Belgique
Tel. (0032) 71 88 56 38

ETUDES UNIVERSITAIRES

1984-1988 **Licence en Sciences - Section Biologie Moléculaire**
Université Catholique de Louvain
1ère Candidature en Sc. Naturelles (Gr. B) (GD)
2ème Candidature en Sciences (Biologie) (GD)
1ère Licence en Sciences (Zoologie) (GD)
2ème Licence en Sciences (Zoologie) (LPGD)

1988-1991 **Diplôme d'Etudes Approfondies en Biologie**
Université Catholique de Louvain (LPGD)

1988-1993 **Doctorat en Sciences - Biologie**
"Transposition et Mécanismes de Spécificité de Cible d'IS*231*A, une Séquence d'Insertion de *Bacillus thuringiensis*"
Université Catholique de Louvain,
Unité de Génétique, prom. prof. Jean Delcour (LPGD+)

EXPERIENCE POSTDOCTORALE ET FONCTIONS ACTUELLES

10/93-10/94 **Assistant de Recherches**
Université Catholique de Louvain, Unité de Génétique prof. Jean Delcour
• Propriétés de transposition des éléments génétiques mobiles IS*231*A et Tn*4430* de *Bacillus thuringiensis*.

10/94-12/97 **Séjour postdoctoral à l'étranger**
University of Oxford, UK, Dpt of Biochemistry, prof. David J. Sherratt
• Mécanisme de recombinaison site-specifique Xer d'*Escherichia coli*.
• Développement d'une méthode de mutagénèse des protéines par insertion aléatoire de pentapeptides à partir du transposon Tn*4430*.

01/98-10/98 **Assistant de Recherches**

10/98-10/01 **Chargé de Recherches FNRS**
Université Catholique de Louvain, Unité de Génétique, prof. Jean Delcour
• Le système de recombinaison site-specifique TnpI/irs et le mécanisme de transposition du transposon Tn*4430*.
• Contrôle génétique de la recombinaison site-spécifique Cre/Lox chez *Lactococcus lactis*.
• Participation aux autres projets du laboratoire sur la biologie moléculaire des bactéries lactiques (métabolisme, synthèse de la paroi, adaptation aux stress et régulation génétique).

10/01- **Chercheur Qualifié FNRS et Chargé de Cours à Temps Partiel à l'UCL**
10/02- **Responsable de l'Unité de Génétique (GENE)**

BOURSES ET CONTRATS

10/88-09/91 **Bourse de thèse IRSIA**

10/91-10/94 **Mandat d'Assistant de Recherches à l'UCL**
(candidature au FNRS reprise par le FDS)
contrat interrompu pour la durée du stage à l'Université d'Oxford

11/94 **Contrat de Recherche à l'Université d'Oxford (MRC)**

12/94-11/96 **Bourse postdoctorale EMBO**

12/96-11/97 **Bourse postdoctorale des Communautés Européennes (BIOTECH)**

12/97-09/98 **Reprise du mandat d'Assistant de Recherches à l'UCL (FDS)**

10/98-09/01 **Mandat de Chargé de Recherches FNRS**

10/01- **Mandat de Chercheur Qualifié FNRS**

DOMAINES D'EXPERTISE

Biologie générale et Moléculaire, Génétique, Microbiologie (Gram$^+$ et Gram$^-$)

Conformation et topologie de l'ADN, structure des protéines, enzymologie

Interactions protéines-ADN, assemblages nucleoprotéiques complexes

Recombinaison génétique et transposition (mécanismes et régulation)

ACTIVITES D'ENSEIGNEMENT (DEPUIS 1998)

Participation aux "Exercices intégrés en Génétique Moléculaire"
(UCL, Biol 2138, J. Delcour - M. Deghorain).
"Modification de la relation structure-fonction de la β-lactamase TEM-1 d' *E. coli* par la méthode de pentapeptidogénèse", 2h-6h/an depuis 1997.

Participation au cours de "Génétique Moléculaire" `
(UCL, Biol 2137, J. Delcour).
"Leçons sur la réplication et les mécanismes de recombinaison génétique", 2-4h/an de 1998-à 2003.

Participation au cours de "Biochimie Comparée"
(UCL, Biol 2271, J. Delcour - R. Rezsöhazy).
"Constance et diversité dans les mécanismes permettant de couper, souder et synthétiser les acides nucléiques", 6h/an de 1999 à 2000.

Chargé de cours invité à l'atelier de DEA sur "la plasticité du génome bactérien" (Univ. P. Sabatier, Toulouse, C. Gutierrez).
"mécanismes et fonctions biologiques de la recombinaison site-spécifique: un pour tous, tous pour un", 4h, novembre 2000.

Co-titulaire du cours de "Biologie moléculaire et cellulaire bactérienne"
(UCL, Biol 2285, J. Delcour - B. Hallet).
"La cellule bactérienne dans le temps", 15h/0h depuis 2001.

Co-titulaire du cours de "Génétique Moléculaire" `
(UCL, Biol 2137, J. Delcour – B. Hallet).
15h/0h depuis 2003.

Co-titulaire du cours de "Génétique Microbienne" `
(UCL, Biol 2211, A. M. Corbisier – B. Hallet).
15h/15h depuis 2003.

VULGARISATION ET PROMOTION DES SCIENCES

Article didactique destiné aux enseignants du secondaire
"Initiation a la démarche biotechnologique"
B. Hallet, (1992), Probio Revue, 15,147-172

Ateliers destinés aux élèves du secondaire
"Du phénotype au gène", 4h/an entre 1990 et 1994.

Conférence au 37ème Congrès Pluraliste des Sciences, Namur, le 25/08/99
"Les transgénoses naturelles et artificielles et leur impact sur l'évolution"
Resp. J. Fraipont, Président de l'Association des Professeurs de Biologie.

Conseiller scientifique de l'exposition "Gène-éthique"
Parc d'Aventure Scientifique "Le PASS" de Frameries.
Resp. C. Bluard, Muséologue au "PASS".

Accueil annuel d'étudiants de l'enseignement secondaire pour l'accomplissement d'un stage de courte durée en entreprise (3 à 5 jours)
(décret de la Communauté Française de Belgique, depuis 2001-2002, 1 à 4 stagiaires par an).

ACTIVITES D'ENCADREMENT (DEPUIS 1998)

PROMOTEUR DE MEMOIRES DE FIN D'ETUDES

V. Vanhooff (1997-1998):	Mise en œuvre de la recombinaison site-specifique Cre-*lox* chez *Lactococcus lactis*
N. Michelet (1998-1999):	Contrôle de la recombinaison site-specifique Cre-*lox* chez *Lactococcus lactis*
C. Galloy (1999-2000):	Construction et caractérisation de sites hybrides entre les systèmes de recombinaison site-spécifique Xer d' *Escherichia coli* et TnpI/*irs* du transposon Tn*4430*
N. Simonis (2000-2001):	Construction d'un système de transposition contrôlée du transposon Tn*4430* de *Bacillus thuringiensis*
M. Lambin (2001-2002):	Caractérisation et purification d'une protéine de fusion entre la transposase de Tn*4430* et un motif polyhistidine
D. Cochonneau (2003-2004)	Construction d'un système hybride pour l'étude des déterminants de l'immunité de transposition de Tn*4430*

PARTICIPATION A L' ENCADREMENT DE MEMOIRES DE FIN D'ETUDES DANS LES AUTRES DOMAINES DE RECHERCHE DU LABORATOIRE

C. Devoghel (1997-1998):	Construction d'une souche de *Lactobacillus helveticus* déficiente en D-lactate déshydrogénase
E. Palumbo (1997-1998):	Construction et caractérisation d'un mutant de L. plantarum déficient pour l'alanine racémase
I. Focant (1998-1999):	Purification et caractérisation de L-lactate déshydrogénase sauvage et mutantes de *Streptococcus thermophilus*
S. Burteau (1998-1999):	Construction d'une souche de *Lactobacillus helveticus* déficiente en D-lactate déshydrogénase et production d'alanine
C. Verheust (1998-1999):	Construction et caractérisation d'un mutant de *Lactobacillus plantarum* déficient pour la glutamate racémase racémase
L. Fontaine (1999-2000):	L'expression chez *L. plantarum* du gène ddl de *L. lactis* subsp. *cremoris* entraîne la sensibilité à la vancomycine et modifie les précurseurs du peptidoglycane
S. Lewahert (1999-2000):	Construction et caractérisation de deux souches mutantes AltA- et AltB- de *Lactococcus lactis* affectées dans le transport de l'alanine
N. Delalune (2000-2001):	Etude des voies de réorientation du pyruvate chez *Streptococcus thermophilus*
W. Glénison (2000-2001):	Etude de la glycosylation des acides teichoiques de parois chez *Lactobacillus plantarum*
A. Grosjean (2000-2001):	Construction de mutants de *Lactococcus lactis* déficients pour le transport de l'alanine
V. Wouters (2000-2001):	Etude de la fonction de la protéine de régulation RelA1 dans la réponse à un choc froid chez *Streptococcus thermophilus*
J-B. Beaudry (2001-2002):	Effets physiologiques de la surexpression des gènes de pyruvate oxydase de *Lactobacillus plantarum*
I. Tytgat (2003-2004):	Importance du D-lactate dans la synthèse de peptidoglycane de *Lactobacillus plantarum*
A. Stulkens (2003-2004):	Etude de la fonction physiologique des pyruvate oxydases de *Lactobacillus plantarum* par surexpression et disruption de gènes *poxD* et *poxF*
E. Verplaets (2003-2004):	Etude du régulateur BlIR du système à deux composantes BlIH/BlIR de *Streptococcus thermophilus*

PROMOTEUR DE THESES DE DOCTORAT

V. Vanhooff, boursière FRIA, depuis sept. (1998-2003)
Le système de recombinaision site-specifique du transposon Tn*4430*
(thèse soutenue le 25/06/2003)

C. Galloy, 3ème année FRIA, depuis sept. 2000:
Interactions protéine-protéines et protéines-ADN impliquées dans l'assemblage du complexe de recombinaison site-spécifique TnpI/*irs* du transposon Tn*4430*

M. Lambin, 2ème année FRIA, depuis sept. 2002:
Etude de la relation structure-fonction de la transposase du transposon réplicatif Tn*4430*

CO-PROMOTEUR DE THESES DE DOCTORAT

S. Leidgens, 1ère année FRIA, depuis sept 2003:
Etude du rôle de la frataxine dans la biosynthèse des centres fer-soufre et de son interaction avec la protéine Isu1P chez la levure *Saccharomyces cerevisiae*
Promoteur F. Foury (Unité FYSA, UCL)

PARTICIPATION A L'ENCADREMENT DE THESES DE DOCTORAT DANS LES AUTRES DOMAINES DE RECHERCHE DU LABORATOIRE

S. Derzelle, boursière FRIA, 1995-2000:
Etude génétique et physiologique de trois gènes *csp* de réponse au choc froid chez *Lactobacillus plantarum*

F. Lorquet, boursière FRIA, 1996-2002:
Etude génétique et fonctionnelle des gènes de la pyruvate oxydase chez *Lactobacillus plantarum*

M. Deghorain, assistante UCL, depuis sept. 1997:
Altération de la synthèse du peptidoglycane de la paroi chez *Lactobacillus plantarum*

P. Goffin, boursier FRIA, depuis sept. 1997:
Approche génétique du rôle physiologique de la racémisation du lactate chez *Lactobacillus plantarum*

E. Palumbo, boursière FRIA, depuis sept. 1998:
Etude génétique de la fonction des substitutions en D-alanine et en glucose des acides téichoïques chez *Lactobacillus plantarum*

L. Fontaine, 2ème année FRIA, depuis sept. 2001:
Etude d'un système de régulation à deux composantes chez *Streptococcus thermophilus*

PARTICIPATION AU COMMITE D'ENCADREMENT DE THESES DE DOCTORAT EFFECTUEES DANS D'AUTRES LABORATOIRES

V. Belis, depuis 2002:
Développement d'une stratégie de sélection pour l'évolution dirigée d'enzymes
Promoteur : P. Soumillion, Unité BIOP, UCL

J. Crouzet, depuis 2002:
Analyse structurelle et fonctionnelle des transporteurs ABC chez la plante
Promoteur : M. Boutry, Unité FYSA, UCL

F. Etienne, depuis 2002:
Etude de la régulation de l'activité de la peroxyrédoxine 5 et de son implication dans la transduction de signaux intracellulaires
Promoteur : B. Knoops, Unité BANI, UCL

D. Vanham, depuis 2002:
Etude du promoteur de transcription de NpABC1, un gène de Nicotiana plumbaginifolia codant pour un transporteur ABC de la membrane plasmique et dont l'expression est induite par le sclareol
Promoteur : M. Boutry, Unité FYSA, UCL

K. Bobik, depuis 2003:
Etude de la régulation de l'ATPase-pompe à propton par phosphorylation
Promoteur : M. Boutry, Unité FYSA, UCL

J. Deherve, depuis 2003
Création d'une activité phosphatase *de novo* par évolution dirigée
Promoteurs : J. Fastrez, P. Soumillion, Unité BIOP, UCL

S. De Simoni, depuis 2003:
Apoptose et nécrose induites par un stress oxydant mitochondrial : étude de la fonction cytoprotectrice de péroxyrédoxine 5 dans la lignée humaine SH-SY5Y
Promoteur : B. Knoops, Unité BANI, UCL

C. Labarbe, depuis 2003 :
Evolution dirigée d'une protéine « PBP like » en beta-lactamase de class A
Promoteurs : J. Fastrez, P. Soumillion, Unité BIOP, UCL

COMMITES DE LECTURE (DEPUIS 1998)

Lecteur de 19 mémoires de fin d'études et de DEA.

Examinateur de 8 thèses de doctorat, dont 5 à l'étranger, et de 2 thèses d'Habilitation à diriger des recherches (HDR)

Evaluation d'articles soumis à plusieurs revues scientifiques (EMBO J., Mol. Microbiol., Enz. Microbiol. Technol., Res. Microbiol.)

COMPOSITION ACTUELLE DE L'EQUIPE RECHERCHE (2002-2003)

Les projets de recherche orientés sur **"les mécanismes moléculaires de transposition et de recombinaison site-spécifique"** de l'Unité de Génétique (2 académique et 14 scientifiques, au total) occupent actuellement **5 chercheurs:**

1 chercheur qualifié:	Bernard Hallet (FNRS)
1 technicienne :	Sylvie Rossenfosse (UCL, 75%)
2 étudiants en thèse de doctorat	Christine Galloy (4ème année FRIA)
	Michaêl Lambin (2ère année FRIA)
1 étudiante en mémoire de licence	Daphné Cochonneau

FINANCEMENT DE LA RECHERCHE (DEPUIS 1998)

10/98-09/00 **Fonds spéciaux de recherche (FSR, UCL):**
24 789 ¤ (fonctionnement)
"Le sytstème de recombinaison site-spécifique du transposon Tn*4430*: étude du mécanisme et applications"

10/00-09/01 **Crédit aux Chercheurs # 1.5.177.01 (FNRS):**
19 831 ¤ (fonctionnement)
"Les interactions moléculaires impliquées dans l'assemblage du complexe de recombinaison et le contrôle de la réaction d'échange des brins d'ADN catalysée par le système de recombinaison site-spécifique Tnpl/*irs* du transposon Tn*4430*"

09/01-10/06 **Action de Recherche Concertée # 01/06-268 (ARC 2000, UCL, Communauté Française de Belgique):**
247 793¤ sur un total de 743 453¤ (personnel, équipement et fonctionnement)
"Accelerated molecular evolution of enzymes" en partenariat avec les laboratoires de J. Fastrez, P. Soumillion et J.-P. Declercq (BIOP/CSTR, UCL)

10/03-10/04 **Crédit aux Chercheurs # 1.5.190.03 (FNRS):**
15 000 ¤ (équipement)
"Les mécanismes moléculaires de la transposition réplicative et l'immunité de transposition du transposon bactérien Tn*4430*: une étude génétique et biochimique"

10/04-10/08 **Fonds de la recherche fondamentale collective # 2.5496.04 (FNRS):**
60 000¤ (équipement et fonctionnement)
«The molecular transposition and site-specific recombination machines of Tn*4430* »

PRINCIPALES COLLABORATIONS SCIENTIFIQUES

Dr D. Lereclus: Institut Pasteur, Paris (depuis 1991)
Etude des éléments génétiques mobiles IS*231* et Tn*4430* de *Bacillus thuringiensis*
(un article en préparation)

Dr D. J. Sherratt: University of Oxford (depuis 1994)
Etude des mécanismes de recombinaison site-spécifique
(stage postdoctoral, 5 articles publiés en commun)

Dr F. Hayes, University of Manchester (depuis 1994)
Développement et utilisation d'une méthode de mutagénèse par insertion aléatoire de peptides basée sur le transposon Tn*4430* ("Pentapeptide Scanning Mutagenesis")
(4 articles publiés en commun)

Drs M. Galleni et J-M Frère, Centre d'Ingéniérie des Protéines, Ulg (depuis 1997)
Utilisation de mutants d'insertion dans la béta-lactamase TEM-1 pour l'expression et la production d'épitopes vaccinaux

Dr J-J Letesson, Unité de Recherche en Biologie Moléculaire, FUNDP (depuis 1998)
Etude de la topologie de la porine OMP2B par insertion aléatoire de peptides ("Pentapeptide Scanning Mutagenesis")

Dr R. Rezsöhazy, Génétique du Développement, UCL (depuis 1998)
Etude de la relation structure-fonction de la protéine homéotique Hoxa-1 par mutagénèse insertionnelle ("Pentapeptide Scanning Mutagenesis")

Dr A. F. Stewart, EMBL, Heidelberg / Max Planck Institute, Dresden (depuis 1998)
Développement et utilisation de la recombinaison site-spécifique TnpI/*irs* chez les mammifères
(un brevet déposé en commun)

Dr S. D. Colloms, University of Glasgow, Glasgow (depuis 1998)
Echange des mécanismes de contrôle entre la recombinaison Xer d'*Escherichia coli* et le système TnpI/IRS de Tn*4430* (un article en préparation)

Dr A. M. Segall, State University of California, San Diego (depuis 2001)
Utilisation de peptides inhibiteurs des topoisomérases et recombinases site-spécifiques
(un article en préparation)

Dr B. Révet, Institut Gustave Roussy, Paris (depuis 2001)
Analyse des complexes nucleoprotéiques par micrsoscopie électronique
(un article en préparation)

Dr P. Soumillion et Prof. J. Fastrez, Unité BIOP, UCL; Prof. J-P Declercq, Unité CSTR, UCL (depuis 2001)
Evolution moléculaire accélérée: développement de nouvelles approches basées sur la recombinase TnpI de Tn*4430*
(Action de Recherche Concertée, ARC)

COMMUNICATIONS SCIENTIFIQUES

18 publications dans des revues internationales avec comité de sélection (liste ci-après)
1 chapitre de livre
3 articles en préparation
1 brevet d'invention
34 communications à des congrès à l'étranger, dont 5 exposés en séance plénière
28 communications à des congrès en Belgique

LISTE BIBLIOGRAPHIQUE

Chapitre de livre

DNA site-specific resolution systems,
Hallet B., V. Vanhooff, and F. Cornet, (2004), p. 145-180 in G. Phillips and B. Funnell (Eds.) *The biology of plasmids.* ASM Press, Washington, D. C.

Publications (revues internationales) **IF 2002**

Improved adaptation to cold-shock, sationary-phase, or freezing stresses in *Lactobacillus plantarum* overproducing cold-shock proteins,
Derzelle S., B. **Hallet**, T. Ferain, J. Delcour and P. Hols (2003)
Appl. Environ. Microbiol., 69,4285-4290. 3.691

Cold shock induction of the *cspL* gene of *Lactobacillus plantarum* involves transcriptional regulation,
Derzelle S., B. **Hallet**, T. Ferain, J. Delcour, and P. Hols (2002),
J. Bacteriol., 184, 5518-5523. 3.959

Playing 'Dr Jekyll & Mr Hyde': combined mechanisms of phase variation in bacteria,
Hallet B., (2001),
Curr. Opin. Microbiol., 4, 570-581. 6.430

Pentapeptide scanning mutagenesis: encouraging old proteins to execute unusual tricks,
Hayes F., .and B. **Hallet**, (2000),
Trends Microbiol., 8, 569-574. 6.665

Changes in *cspL*, *cspP*, and *cspC* mRNA abundance as a function of cold shock and growth phase in *Lactobacillus plantarum*,
Derzelle S., B. **Hallet**, K. P. Francis, T. Ferain, J. Delcour, and P. Hols, (2000),
J. Bacteriol. 182, 5105-5113. 3.959

Coupled catalysis in a recombination machine: Mutated XerC and XerD recombinases that stimulate strand exchange by their partner recombinase,
Arciszewska L. K., R. Baker, B. **Hallet**, and D. J. Sherratt, (2000),
J. Mol. Biol. 299, 391-403. 5.359

Reciprocal control of catalysis by the tyrosine recombinases XerC and XerD: an enzymatic switch in site-specific recombination,
Hallet B., L. K. Arciszewska, and D. J. Sherratt, (1999),
Mol. Cell 4, 949-959. 16.471

Structure-Function Correlations in the XerD Site-Specific Recombinase Revealed by Pentapeptide Scanning Mutagenesis,
Cao Y., B. **Hallet**, D. J. Sherratt, and F. Hayes, (1997),
J. Mol. Biol. 274,39-53. 5.359

Transposition and site-specific recombination: adapting DNA cut and paste mechanisms to a variety of genetic rearrangements,
Hallet B., and D. J. Sherratt, (1997),
FEMS Microbiol. Rev. 21, 157-178. 9.597

Insertion mutagenesis as a tool in the modification of protein function: extended substrate specificity conferred by pentapeptide insertions in the Ω loop of TEM-1 β–lactamase,
Hayes F., B. **Hallet**, and Y. H. Cao, (1997),
J. Biol. Chem. 272, 28833-28836. 6.696

Xer recombination in *Escherichia coli*. Site-specific DNA topoisomerase activity of the XerC and XerD recombinases,
Cornet F., B. **Hallet**, and D. J. Sherratt, (1997),
J. Biol. Chem. 272, 21927-21931. 6.696

Pentapeptide scanning mutagenesis: random insertion of a variable five amino acid cassette in a target protein,
Hallet B., D. J. Sherratt, and F. Hayes, (1997),
Nucleic Acids Res. 25, 1866-1867. 7.051

IS*231*A insertion specificity: consensus sequence and DNA bending at the target site,
Hallet B., R. Rezsöhazy, J. Mahillon, and J. Delcour, (1994),
Mol. Microbiol. 14, 131-139. 5.832

IS*231* and other *Bacillus thuringiensis* transposable elements: a review,
Mahillon J., R. Rezsöhazy, B. **Hallet**, and J. Delcour, (1994),
Genetica. 93, 13-26. 1.063

The IS*4* family of insertion sequences: evidence for a conserved transposase motif,
Rezsöhazy R., B. **Hallet**, J. Delcour, and J. Mahillon, (1993),
Mol. Microbiol. 9, 1283-1295. 5.832

IS*231*V and W from *Bacillus thuringiensis* subsp. *israelensis*, two distant members of the IS*231* family of insertion sequences,
Rezsöhazy R., B. **Hallet**, J. Mahillon, and J. Delcour, (1993),
Plasmid. 30, 141-149. 1.495

IS*231*D, E and F, three new insertion sequences in *Bacillus thuringiensis*: extension of the IS*231* family,
Rezsöhazy R., B. **Hallet**, and J. Delcour, (1992),
Mol. Microbiol. 6, 1959-1967. 5.832

IS*231*A from *Bacillus thuringiensis* is functional in *Escherichia coli*: transposition and insertion specificity,"
Hallet B., R. Rezsöhazy, and J. Delcour, (1991),
J. Bacteriol. 173, 4526-4529. 3.959

Articles en préparation

Self-control in DNA recombination mediated by the tyrosine recombinase TnpI,
Vanhooff V., C. Galloy, H. Agaisse, D. Lereclus, B. Révet, and B. **Hallet**
Manuscrit prêt pour soumission à EMBO J.

Relative contribution of the recombination site core and accessory sequences in the directionality of the site-specific recombination reaction catalysed by the tyrosine recombinase TnpI,
Vanhooff V., C. Normand, C. Galloy, A. M. Segall, and B. **Hallet.**

Exchanging control mechanisms between DNA site-specific recombination systems of the tyrosine recombinase family,
Galloy C., V. Vanhooff V., C. Normand, S. D. Colloms, and B. **Hallet.**

Brevets d'invention

A new tyrosine recombinase for genetic engineering,
Stewart A. F., Y. Zhang, and B. **Hallet**, (2001),
US Patent Application # 09/895.435, United States of America

Annu. Rev. Biophys. Biomol. Struct. 2001. 30:87–104

A STRUCTURAL VIEW OF Cre-*loxP* SITE-SPECIFIC RECOMBINATION

Gregory D. Van Duyne
Department of Biochemistry and Biophysics, Howard Hughes Medical Institute, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104; e-mail: vanduyne@mail.med.upenn.edu

Key Words recombinase, integrase, synapse, Holliday junction, strand exchange

■ **Abstract** Structural models of site-specific recombinases from the lambda integrase family of enzymes have in the last four years provided an important new perspective on the three-dimensional nature of the recombination pathway. Members of this family, which include the bacteriophage P1 Cre recombinase, bacteriophage lambda integrase, the yeast Flp recombinase, and the bacterial XerCD recombinases, exchange strands between DNA substrates in a stepwise process. One pair of strands is exchanged to form a Holliday junction intermediate, and the second pair of strands is exchanged during resolution of the junction to products. Crystal structures of reaction intermediates in the Cre-*loxP* site-specific recombination system, together with recent biochemical studies in the field, support a "strand swapping" model for recombination that does not require branch migration of the Holliday junction intermediate in order to test homology between recombining sites.

INTRODUCTION

Site-specific recombinases from the lambda integrase family of enzymes catalyze DNA rearrangements that are critical for a variety of important biological functions. Perhaps the best known example is the integration and excision of the bacteriophage λ genome into and out of the *Escherichia coli* host chromosome (9). Other functions include the resolution of multimeric plasmids and chromosomes to monomers in order to ensure faithful segregation upon cell division, the amplification of yeast 2μ circle copy number, and the regulation of gene expression (reviewed in 24, 44, 57, 59). The λ-integrase family, also referred to as the "tyrosine recombinases" (58), includes over 100 members identified on the basis of sequence similarity (47). The most well-studied examples include, in addition to the integrase protein from bacteriophage λ (34), the bacterial XerC and XerD recombinases (57), Cre recombinase from bacteriophage P1 (1), and the Flp recombinase from *Saccharomyces cerevisiae* (53).

1056-8700/01/0610-0087$14.00

The tyrosine recombinases carry out the site-specific recombination reaction in a stepwise manner. One pair of DNA strands is first exchanged to form a Holliday junction intermediate, and then the Holliday junction is converted to recombinant products by exchange of the second pair of DNA strands. In contrast to the many years of biochemical and molecular genetic studies on the various integrase family members by several laboratories, structural models for the tyrosine recombinases have only recently become available, beginning with the report of the λ-integrase catalytic domain crystal structure in early 1997 (33). Since then, structures of the HP1 integrase catalytic domain (27), the XerD protein (62), and three Cre recombinase/DNA complexes (17, 22, 23) have been described. Most recently, the structure of an Flp recombinase/DNA complex has been determined (11a). The resulting structural models for the tyrosine recombinase family have provided both a view of the enzyme active sites and a three-dimensional perspective on the reaction pathway (for minireviews of the structural work, see references 20, 36, 40, 58, 68). This review focuses on the three-dimensional nature of the Cre-*loxP* site-specific recombination pathway, where models of the recombinase-DNA complexes for three intermediates in the reaction are available.

THE Cre-*loxP* SYSTEM

Cre recombinase is a 38-kDa protein encoded by bacteriophage P1. Its roles in the P1 life cycle are thought to include cyclization of the linear genome and resolution of dimeric chromosomes formed following DNA replication as a result of homologous recombination (61). The DNA sequences where recombinase binding and strand exchange take place are named *loxP* in the Cre recombinase system. These sites are composed of two recombinase binding elements (RBEs) arranged as nearly perfect inverted repeats surrounding a central strand exchange or crossover region (Figure 1). Two recombinase subunits bind to each core site (one to each

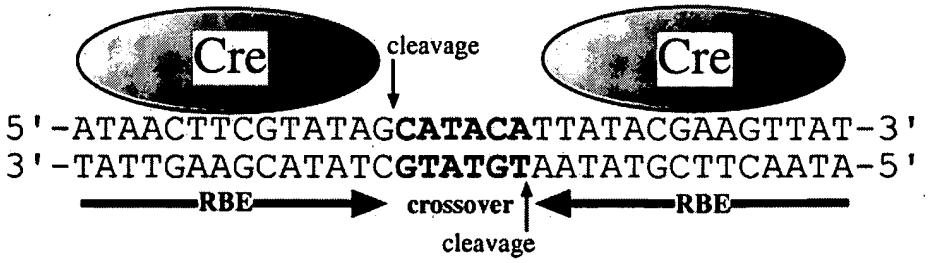


Figure 1 The 34-bp *loxP* site is composed of two 14 bp recombinase binding elements (RBEs) arranged as inverted repeats around a central 6-bp crossover region (28). Cleavage of the sites occurs at the borders between the crossover region (in boldface) and the RBEs. Two Cre recombinase subunits bind cooperatively to the loxP site (one to each RBE) with no direct contact between the recombinases and bases in the crossover region.

RBE) with a high level of cooperativity (7, 51). The phosphoryl transfer strand exchange chemistry between substrates occurs within the 6-bp crossover region. The crossover sequence is asymmetric and therefore provides directionality to the site. For the simplest systems, exemplified by Cre recombinase, this core recombination site is sufficient for the complete recombination reaction (3, 53). For more complex systems, such as λ-integrase and the XerC/D recombinases acting on plasmid substrates, the sites required for recombination contain additional sequences recognized by accessory proteins and auxiliary DNA-binding domains. The simplicity of the Cre recombinase system has led to its widespread use as a tool in the manipulation of DNA molecules both in vivo and in vitro (54).

The tyrosine recombinases use a topoisomerase I-like mechanism to cleave and religate DNA strands during recombination. An historical model for the recombination pathway is shown in Figure 2, which illustrates the stepwise nature of the reaction. This view of the reaction involves only the core recombination sites and the core-binding components of the recombinase enzymes. For simple systems (such as Cre) recombining linear substrates, this model represents all of the interacting components. For more complex systems, there are additional factors that play architectural and/or energetic roles that could influence the various steps in the recombination pathway. For reviews of the λ-integrase family recombination mechanism as discussed prior to 1995, see references (13, 34, 35, 44, 59).

In the mechanism shown in Figure 2, two Cre-bound *loxP* sites associate to form a recombination synapse. Two of the four recombinase subunits in the synapse cleave the DNA substrates using conserved tyrosine side chains as nucleophiles, forming covalent 3′-phosphotyrosine linkages to DNA and releasing free 5′-hydroxyl groups. Either the 5′-hydroxyl groups can reseal the nicks to restore the original substrates and complete one round of topoisomerase-I-like cleavage and religation, or the 5′-hydroxyl groups can attack the 3′-phosphotyrosine linkages of the partner substrates, resulting in the exchange of one pair of strands and formation of a Holliday junction (HJ) intermediate. After branch migration through the crossover region, the second pair of strands in the HJ intermediate are then cleaved and exchanged by the second pair of recombinase subunits to form recombinant products.

The tyrosine recombinases display a strong requirement for sequence identity in the crossover region between recombining sites. Early models have explained how each site senses homology with its partner by requiring the HJ intermediate to branch migrate through this region (66). Mismatches in sequence would provide energetic barriers to this process and ensure that only identical crossover regions could be efficiently recombined. The branch migration model is satisfying in terms of explaining the requirements for homology and the observed reversibility of the recombination pathway, but it is difficult to reconcile with the motions required of the recombinase subunits and DNA arms of the HJ in three dimensions (60). In order to branch migrate 6-8 bp, the recombinase-bound arms of the HJ intermediate would need to undergo extensive rotations, requiring that protein-protein interfaces formed in the synapse be disrupted. In the mid 1990s, Landy and coworkers

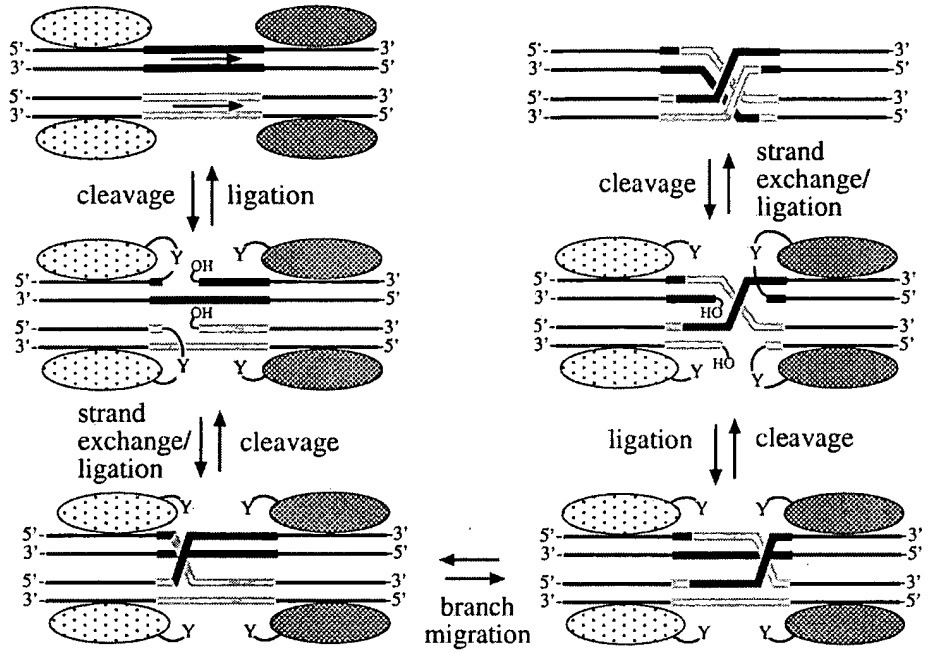


Figure 2 The branch migration model for integrase family site-specific recombination (for a review of this mechanism, see reference 13). Two recombinase-bound sites associate to form a recombination synapse (*top left*). Two subunits cleave the DNA substrates with conserved tyrosine side chains to form 3′-phosphotyrosine linkages and release free 5′-hydroxyl groups (*middle left*). The 5′-hydroxyl groups undergo intermolecular attack of the partner phosphytyrosine to complete the exchange of one pair of DNA strands between the two substrates and form a Holliday junction intermediate (*lower left*). The branch point of the junction starts at the site of initial strand exchange and then migrates through the crossover region to the second set of cleavage sites. The second pair of subunits is then activated and/or positioned for cleavage of the bottom substrate strands, which are exchanged to form recombinant products. Heterology between crossover sequences would block efficient branch migration and prevent the reaction from proceeding to the second strand exchange. For simplicity, the DNA sites are shown associating in a parallel orientation in this figure. The same mechanism can also be described with an antiparallel alignment of sites (see reference 4).

proposed an alternative "strand swapping-isomerization" model for integrase family site-specific recombination (45) that was supported by experimental results from a number of laboratories reported at nearly the same time (4, 8, 37, 69). In this mechanism, strand exchange occurs following cleavage of the site by melting 2–3 bases from their complementary strand and annealing to the corresponding complementary strand in the recombining partner. Those bases are then effectively tested for homology with the partner by Watson-Crick base-pairing and subsequent ligation. As discussed in this review, the Cre-DNA structural models strongly

support the strand swapping, but not the branch migration model for integrase family site-specific recombination.

THE STRUCTURE OF Cre RECOMBINASE

Before three-dimensional structural data became available, most of what was known about protein structures in the tyrosine recombinase family came from sequence comparisons (2, 6) and limited proteolysis experiments (29, 48). The tyrosine recombinases share limited sequence similarity overall, although it is now clear that family members have a common catalytic domain fold (15, 47). Early sequence analyses revealed four strictly conserved amino acids required for catalysis: the Arg-His-Arg catalytic triad and the tyrosine nucleophile (2, 6). All four residues were found to be in the C-terminal halves of the recombinase sequences, which show higher sequence similarity than do the more divergent N-terminal halves. As illustrated in Figure 3, Cre is a two-domain protein with a helical N-terminal domain and a larger, primarily helical C-terminal domain.

The complexes of Cre recombinase with DNA substrates revealed an extensive protein-DNA interface between each protein subunit and its contacted RBE (17, 22, 23). The two domains of Cre form a C-shaped clamp that grasps the DNA from opposite sides, where the N-terminal domain interacts primarily with the major groove proximal to the crossover region on one face, and the C-terminal domain interacts with successive minor, major, and minor grooves on the opposite face (Figure 3a). A similar organization has been observed in the Flp-DNA complex (11a) despite clear differences in the structures of the proteins N-terminal to the conserved core catalytic domains. Biochemical evidence in the Xer and λ-int systems, together with the available structural models, indicates that this recombinase-DNA organization is likely to be a common feature of the tyrosine recombinases (26, 63).

In addition to the extensive protein-DNA interface formed between Cre and its recombination site *loxP*, a substantial protein-protein interface is formed between the two subunits bound to RBEs on the same site. On one side of the DNA substrate, the two N-terminal domains interact with one another primarily through helix-helix contacts, but they do not interact substantially with the catalytic domains (Figure 4a). On the opposite side, the two catalytic domains interact via the exchange of helices located at the carboxyl termini (Figure 4b). The C-terminal helix in one Cre subunit (helix-N) buries its hydrophobic surface in an acceptor pocket on the adjacent subunit. The network of contacts between N- and C-terminal domains in DNA-bound Cre subunits may in part explain the large cooperativity of binding to the *loxP* site (1, 51).

The C-terminal helix exchange is not reciprocal between Cre subunits bound to the same *loxP* site in the Cre-DNA crystal structures because a second Cre-*lox* complex interacts to form a synaptic assembly and the helix exchange is cyclic among the four subunits (Figure 4b). In the HP1-integrase catalytic domain struc-

ture, however, a C-terminal helix swap occurs in a reciprocal fashion, leading to an HP1-integrase dimer (27). An HP1-like reciprocal exchange of C-terminal helices between DNA-bound Cre subunits appears to be precluded for geometric reasons if the *loxP* site is bent even slightly. Modeling exercises indicate that even in the case of an unbent *loxP* site, mutual exchange of C-terminal helices would be difficult and may require partial unfolding or repositioning of the penultimate helix (M) in order to form the correct helix-acceptor interactions (21).

The helix-swapping arrangement observed in the Cre-DNA structures not only provides a model for understanding cooperative binding, bending, and synapsis of *loxP* sites, but it also provides a provocative mechanistic argument for control of the cleavage reaction. The conserved tyrosine nucleophile (Tyr324) is located close to the peptide linker (Arg326-Gly333) that connects helix-N to the rest of the domain. Changes in quaternary structure of the recombination complex could therefore be sensed by the tyrosine nucleophile through this peptide linker, providing a stereochemical coupling between the overall structure of the recombination assembly and the positioning of a critical catalytic amino acid.

SYNAPSIS AND STRAND EXCHANGE

In order for recombination to occur, two DNA sites must associate to form a synaptic complex, within which the cleavage and strand exchange reactions take place. In the structures of two Cre mutants bound to symmetrized *loxP* sites (22), the Cre-bound DNA duplexes are bent sharply within the crossover regions and are brought together with all four arms (defined as the DNA segments on either sides of the bends) lying nearly in the same plane (Figure 4). The four recombinase monomers create a pseudo fourfold-symmetric network of protein-protein interactions responsible for holding the synapse together, with a set of interactions between the N-terminal domains on one side of the DNA plane and an independent set of interactions between C-terminal domains on the opposite side of the DNA substrate plane. The complex is strictly twofold symmetric overall, with the dyad of symmetry passing through the center of the complex in a direction perpendicular to the plane of the DNA substrates. A similar twofold-symmetric arrangement of recombinase and substrate is present in the structure of the cleaved covalent Cre-DNA intermediate (23) and in the structures of the Cre- and Flp-bound Holliday junction intermediates (11a, 17). Although the DNA sites in these cases are symmetric and therefore have no sequence-derived directionality that would specify parallel versus antiparallel alignment, the twofold-symmetric structures strongly imply an antiparallel alignment of wild-type recombination sites. The three-dimensional mechanism discussed for the Cre-*loxP* pathway (18, 22) in fact requires that the sites be aligned in an antiparallel fashion in order for productive strand exchange to occur in the absence of large structural rearrangements.

The nature of the protein-protein contacts in the Cre-DNA synaptic complexes, and more recently in the Flp-HJ complex, have provided a great deal of insight into

many aspects of the recombination process. A striking feature of these structures is the types and the extents of the protein-protein interfaces. In the Cre tetramer, the self-association of N-terminal domains alone buries ~4000 Å^2 of solvent accessible surface area, with the interactions between C-terminal domains burying a further ~7000 Å^2. The N-terminal domains primarily interact with one another (and with the DNA) to form a nearly fourfold-symmetric tetramer via helix-helix and loop-helix interactions. As a consequence of this symmetry, the interactions between two N-terminal domains bound to the same *loxP* site are nearly identical to the synaptic interactions between N-terminal domains on different sites. This arrangement of N-terminal domains is capable of accommodating the synapsed substrates, the cleaved substrates, and the Holliday junction intermediate in both halves of the pathway with virtually no changes in quaternary structure. The N-terminal domains therefore appear to form a relatively rigid structure that persists throughout the reaction.

Some of the most informative insights from structural models in the tyrosine recombinase system concern the geometry of the cleavage and strand exchange steps in the reaction. Our first snapshot of this process in action came from the crystal structure of Cre recombinase covalently bound to a suicide DNA substrate (23). This intermediate was formed using symmetrized *loxP* sites that contained nicks adjacent to the scissile phosphates. Cleavage by the recombinase proteins resulted in formation of 3′-phosphotyrosine linkages to the DNA and release of free cytidine molecules, which diffuse away from the active sites. The resulting complex was trapped at the strand exchange step because the departing cytidine takes with it the 5′-hydroxyl required to either continue or reverse the strand exchange reaction, and the remaining strand is too short on the 5′-end to reach either of the activated phosphotyrosines.

The covalent Cre-DNA complex revealed a synaptic assembly with architecture nearly identical to that later observed in the precleavage synaptic complex (Figure 5). The recombinase proteins form a pseudo-fourfold-symmetric tetramer with the DNA duplex substrates in a distorted square planar arrangement. Two of the four protein subunits diagonally related in the complex have cleaved the DNA to form covalent bonds to the scissile phosphate, but the other two subunits have not cleaved the DNA. Remarkably, the 5′-ends of the DNA strands that would be transferred to the partner substrates in the normal reaction (if the 5′-cytidines were present) have partially melted away from their complementary strands, with their 5′-thymidine ends converging in the middle of the synaptic complex (see Figure 5b).

It is tempting to think of this cleaved intermediate structure as a snapshot of the strand exchange process in action. In this context, the complex structure could equally well represent a trapped strand transfer reaction in the HJ-forming step or in the HJ-resolving step. One of the interesting questions concerning this view of the structural model is whether the observed protein-DNA architecture needs to undergo a significant change in quaternary structure in order to accommodate the transfer and ligation of DNA strands between opposing halves of the recombination

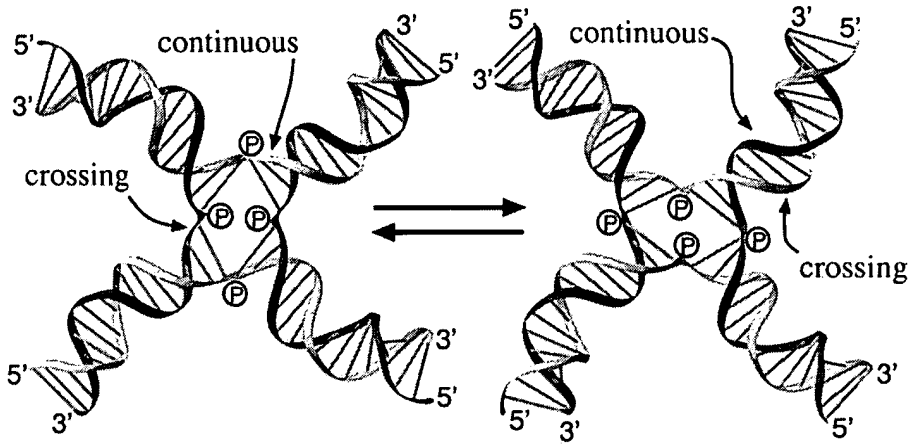


Figure 6 Isomerization model for the HJ intermediate in Cre-*loxP* site-specific recombination. The conformer on the left differs from that on the right by an exchange in the interarm angles and an exchange in the positions of the branch-point phosphates. The stereochemical identities of the dark (crossing) strands on the left are identical to those of the light (crossing) strands on the right. Likewise, the light strands (continuous) on the left are equivalent to the dark (continuous) strands on the right. The strands labeled "crossing" are activated for cleavage and exchange in the two conformers.

assembly. Modeling exercises suggest that the intact strands (with the 5′-cytidine restored) would be able to reach the complementary strands of the partner substrate and form at least two Watson-Crick base pairs immediately adjacent to the scissile phosphate. This annealing process could occur by moving only the three nucleotides at the 5′-ends of the DNA strands and leaving the remainder of the protein-DNA assembly unchanged. With more extensive movement of the DNA strands, a third base pair (and thus, a fully base-paired HJ intermediate) could be formed. The structure of the Cre-HJ intermediate (discussed below) indicates that three bases are exchanged between duplex substrates, placing the branch point of the HJ formed at the center of the *loxP* site. This implies that exactly half of the 6-bp crossover site between scissile phosphates are exchanged in forming the HJ (Figure 6) and the remaining half are exchanged in the resolution of the HJ intermediate to products.

THE ORDER OF STRAND EXCHANGE

In principle, either of the DNA strands in the two duplex substrates could be cleaved and exchanged first in the reaction pathway. From a structural point of view, it is useful to consider whether there are differences between the DNA strands and/or the active sites in the precleavage synaptic complex that would favor the initial

cleavage and exchange of one strand over the other. In the Cre-DNA synapse formed with inactive Cre mutants and symmetrized *loxP* sites (22), there is a clear distinction between the two strands. The strand that corresponds to the one caught in the act of swapping has a helicoidal trajectory that takes it through the middle of the open strand exchange region between duplex substrates. This strand is also relatively free from recombinase-backbone contacts that might limit its mobility. In contrast, the complementary strand is embedded in a more extensive protein interface, which is primarily composed of sugar and phosphate contacts located on the side of the DNA duplex pointing away from the strand exchange region. If this strand were to be exchanged first, a more complicated unraveling from the duplex and dissociation from recombinase interactions would be required.

These observations from the precleavage Cre-DNA complex are consistent with the DNA strand that was caught in the act of exchange in the covalent intermediate structure. Together they suggest that the order of strand exchanges in the recombination pathway is determined once the synaptic complex is formed. However, two distinct antiparallel (and two distinct parallel) synapses are possible, which differ by the direction of bending of the sites. Recombination would therefore be expected to proceed with an opposite order of strand exchanges in these two cases (22). For Cre, a preference for cleaving at the "G/C" end of the *loxP* crossover (left side in Figure 1) has been reported (30), which indicates that formation of one of the two antiparallel synapses is slightly favored over the other. In principle, a bending preference arising from the crossover sequence itself could be responsible for the observed bias, although there are no published experimental data to support this idea for the Cre-*loxP* system. Independent biochemical evidence based on studies of bulged DNA substrates in the Cre and Flp systems also supports this relationship between bending of the core recombination site and strand cleavage preference in the synaptic complex (38).

The Flp recombinase, which has a larger crossover region compared to Cre, shows little or no preference for the order of strand exchange during recombination (38, 39). Although bending of the *frt* site by Flp has been well studied (42, 43), there is no strong evidence that the bend occurs in a preferred direction, which is consistent with a lack of initial cleavage preference. The λ-Int and Xer systems, on the other hand, show a clear preference for cleaving one strand first during recombination (5, 31, 46).

THE HOLLIDAY INTERMEDIATE

At present, there are two crystal structures available representing the Cre-HJ intermediate (17) and, more recently, a crystal structure of Flp recombinase bound to an HJ substrate with a 7-bp crossover site (11a). In one of the Cre-bound junctions (HJ2 in reference 16), eight overlapping DNA strands are assembled to form an HJ with four nicks (missing phosphates) that is fourfold symmetric with respect to sequence symmetry and in principle is free to branch migrate within the crossover

region. The branch point is located at the center of the crossover region in this structure. The other Cre-bound HJ (HJ1 in reference 17) is an immobile junction with the branch point fixed by design at the center of the crossover region. The structures of the two junctions are nearly superimposable, indicating that the presence of the nicks and the asymmetries imposed to form an immobile junction most likely do not distort the observed structures significantly from that of the true HJ intermediate in Cre-*loxP* site-specific recombination.

The Cre-bound HJ arms, defined here as the duplex DNA segments extending away from the branch point, lie nearly in the same plane with a small amount of curvature within each arm that creates a slightly concave surface on the side of the junction plane where the recombinase catalytic domains bind. The arms adopt a distorted square planar arrangment that is strictly only twofold symmetric, with interarm angles of ~75° and ~105° (Figure 7). At the center of the junction all bases are Watson-Crick paired, and all eight bases are unstacked and exposed to solvent. The nearly square planar arms and the unstacking at the branch point in the Cre-HJ complexes most closely resemble the square planar form of the HJ characterized in solution by Lilley and coworkers (14) and observed in the crystal structures with bound RuvA protein (25, 52). This free junction conformer exists in the absence of divalent ions or high concentrations of monovalent ions, where phosphate-phosphate repulsions presumably lead to a fully extended fourfold symmetric structure (41).

In addition to the unequal interarm angles, the geometry of the deoxyribose-phosphate backbone is quite different for the two pairs of DNA strands in the Cre-bound HJ. The strands that span the obtuse interarm angle form a more or less uninterrupted helicoidal trajectory through the branch point of the junction. The DNA strands that span the acute interarm angles have a rather sharp discontinuity at the branch point, where the branch point phosphate rotates toward the center of the junction, away from the path traced by the flanking phosphates (Figure 7b). Thus, in the Cre-bound HJ, like in the stacked-X form of the junction in solution (14), there are a pair of continuous DNA strands and a pair of crossing DNA strands that are stereochemically distinct. It follows, therefore, that the recombinase active sites that surround the scissile phosphates on the continuous versus crossing strands in the junction are not stereochemically equivalent. Based on a comparison with the covalent intermediate complex (Figure 5), the crossing strand active sites are activated for cleavage and the continuous strand active sites are prevented from cleaving the DNA during recombination.

It has been understood for many years that a change in the structure of the HJ intermediate in tyrosine recombinase recombination must serve the pivotal role of deciding whether the reaction should proceed forward to recombinant products or backward to restore the original substrates (see discussion in references 13, 17, 45). Early models of this isomerization step involved a migration of the junction branch point between the sites of cleavage within the crossover region (Figure 2). The strands cleaved in resolution of the HJ would be those whose

scissile phosphates were nearest the junction branch point in a given branch-point isomer. In addition to providing a mechanism for reading homology between sites, this model provided a rationale for how pairs of active sites could be turned on and off in response to the quaternary structure of the recombination complex. The Cre-DNA complex structures, and in particular the structures of the Cre-HJ intermediates, have provided a different view of this central isomerization step in the recombination pathway, which along with biochemical studies from several laboratories has led to an alternative model (17, 18).

In the specific case of the Cre-bound HJ complex, there are two different ways one could draw the structure shown in Figure 7b. The conformers differ by which arms form acute and which arms form obtuse angles, and by which DNA strands adopt the crossing configuration with inverted phosphates at the branch point and which strands adopt the more continuous configuration. The Cre-loxP model for HJ isomerization shown in Figure 6 involves a planar scissoring motion of the HJ arms that swap the pairs of interarm angles along with adjustments to the deoxyribose-phosphate backbone torsion angles that lead to a swap of branch point phosphate configurations. The analogy with crossover isomerization of the Holliday intermediate in homologous recombination is evident. In both cases, the generation of strand equivalence by quaternary changes in the junction structure is accomplished.

The effect of HJ isomerization on the catalytic domains of the recombinase subunits mirrors that of the DNA strands. Those pairs of adjacent subunits that are separated by a larger distance in one conformer are closer together in the alternative conformer. Accordingly, the nature of the interactions between the catalytic domains in these two distinct interfaces is exchanged between recombinase pairs. These differences in domain-domain interfaces may directly relate to the positioning of active site residues and therefore in determining which subunits are activated for cleavage. The situation with the N-terminal domains of the Cre subunits is somewhat different. These domains maintain a nearly fourfold-symmetric arrangement in both isomers, although the interactions formed with the crossing and continuous strands of the junction are slightly different.

In terms of active versus inactive subunits in the Cre-HJ model, the subunits whose active sites are positioned on the crossing strands are predicted to be active for strand exchange, and the pair of subunits on the continuous strands are predicted to be rendered inactive. These predictions follow from a comparison with the cleaved Cre-DNA intermediate (23), where the active site that has catalyzed formation of a covalent Cre-DNA linkage corresponds to one containing the crossing strand phosphates in the HJ complex. A similarity also exists at the quaternary structural level between the synapsed substrates and the HJ intermediate. Aside from differences in DNA structure at the center of the synaptic complex due to the strand exchange event, the structures of the synapsed sites, the covalent Cre-DNA intermediate, and the Cre-HJ intermediate are nearly superimposable, allowing for easy prediction of the active subunits in the two complexes where direct evidence of cleavage is not available.

It is clear from Figure 6 that the proposed isomerization of the Cre-HJ intermediate does not involve migration of the branch point from the center of the crossover region. In fact, given the extent of the protein-protein interfaces in the Cre tetramer (average of ~3400 Å^2 surface buried in each pair of adjacent subunits), it is difficult to imagine how this HJ intermediate could undergo any branch migration. A coordinated rotation of the protein-bound DNA arms to move the branch point would require that the observed protein-protein interface be disrupted and a new interface be formed. Even if the old and new interfaces were isoenergetic, considerable activation energy would no doubt be required to accomplish this type of rearrangement.

RECOMBINASE ACTIVE SITES AND PHOSPHORYL TRANSFER

The active site model that has emerged from the integrase family protein and protein domain structures (27, 33, 62), the closely related eukaryotic topoisomerase Ib structures (12, 49), and an analysis of three intermediates in the Cre-DNA pathway is illustrated in Figure 8. The previously identified Arg-His-Arg side chains coordinate the scissile phosphate during recombination, while the conserved tyrosine side chain is positioned near the scissile phosphate, poised for nucleophilic attack. In addition to these four residues, two new participants in the enzyme active

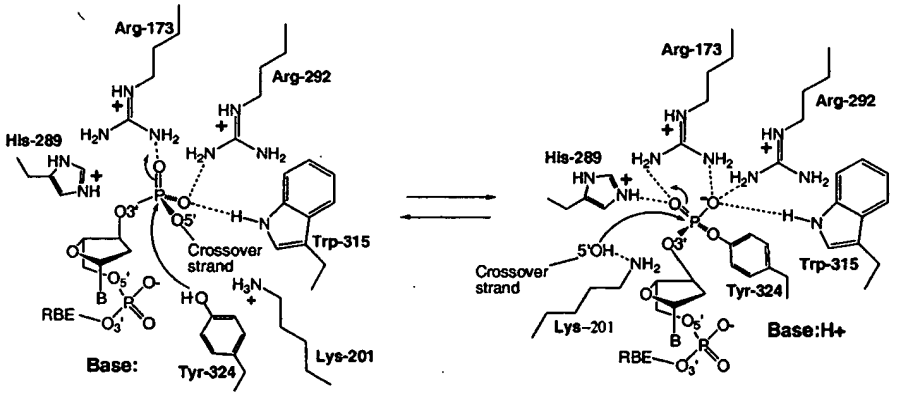


Figure 8 Active site cleavage model for Cre recombinase. The general base responsible for accepting a proton from the tyrosine nucleophile during cleavage has not been clearly identified biochemically but could be His-289. The general acid responsible for protonating the 5′-OH leaving group has been identified as the equivalent of Lys-201 in the vaccinia virus topoisomerase and is therefore likely to play the same role in the tyrosine recombinases, as shown here (32). The ligation reaction is simply the reverse of that shown above, beginning with nucleophilic attack of the 3′-phosphotyrosine intermediate by a 5′-hydroxyl group.

site have been identified. A tryptophan forms a hydrogen bond to the scissile phosphate via the indole nitrogen in Cre, whereas a histidine side chain occupies this position in the sequences of λ-Int, HP1-Int, XerD, and the vaccinia and human topoisomerases. A comparison of tyrosine recombinase sequences indicates that this His/Trp hydrogen bond is likely to be conserved throughout the protein family.

The second previously unidentified active site component is a lysine residue located in the loop between β-strands 2 and 3 (Lys201 in Cre). Structural and biochemical studies in the vaccinia virus topoisomerase system were key to realizing the importance of this residue in the site-specific recombinases (12, 67). In the vaccinia virus enzyme, mutation of the corresponding residue (Lys165) results in a loss of 10^4 in catalytic efficiency (67). The alanine mutants of this lysine in XerD (10) and Cre (F Guo & G Van Duyne, unpublished observations) are each defective in recombination. In the covalent Cre-DNA and human topoiosomerase-Ib DNA structures, this lysine contacts a base adjacent to the cleaved phosphate in the minor groove (23, 49). In the synaptic Cre/DNA complexes and in the Cre/HJ complex structures, the loop containing Lys-201 is not well ordered, most likely indicating a high degree of mobility.

A general model for acid-base catalysis of the initial phosphoryl transfer step in the tyrosine recombinase cleavage reaction is shown in Figure 8 (see also references 19 and 58). The Arg-His-Arg triad of conserved side chains, along with the Trp side chain observed in Cre, provides an array of hydrogen bond donors surrounding the scissile phosphate that fulfills the requirement of stabilizing a pentacoordinate transition state of the phosphate and probably contributes an electrophilic catalysis component to the reaction. The equivalent of Lysine-201 has been shown to be the catalytic acid responsible for protonating the 5′-OH leaving group in the vaccinia virus topoisomerase system (32), and it seems quite likely that the same is true in the tyrosine recombinases. The identity of a general base that accepts the Tyr-324 hydroxyl proton is still uncertain, although His-289 is a prime candidate.

SUMMARY

The convergence of biochemical, genetic, and structural studies in an area of biological research generally results in a strengthened view of the underlying molecular processes. This is clearly the case for the tyrosine recombinase family of site-specific recombination enzymes. Structural data for five of the integrase family recombinase systems and two eukaryotic topoisomerase Ib systems have provided a three-dimensional framework for understanding years of results obtained by a number of different laboratories. Biochemical studies in the vaccinia virus topoisomerase Ib enzyme have also contributed greatly to our understanding of catalysis in this superfamily.

Our current understanding of the tyrosine recombinase site-specific recombination pathway in the three-dimensional sense has been guided primarily by structures of reaction intermediates in the Cre-loxP system. The most surprising

feature of what we have learned from this system is that the reaction appears to proceed with a single protein-DNA architecture that requires no large changes in quaternary structure in converting substrates to products. A subtle isomerization of the nearly coplanar Holliday junction intermediate between dyad-symmetric forms may be sufficient to act as the active site switch that triggers the exchange of only one of the two pairs of DNA strands to generate either substrates or products. The current model for the Cre-*loxP* recombination pathway is outlined in Figure 9. This mechanism requires antiparallel synapsis of the sites in order to lead to productive strand exchange. The strand exchange process occurs by a strand-swapping mechanism based on that proposed in 1995 by Nunes-Düby et al (45). Three bases are tested for homology with the recombining site and exchanged to form the Holliday intermediate, and the remaining three bases in the crossover region are tested during the second strand exchange step.

Perhaps the most compelling questions remaining at the structural and mechanistic level for the tyrosine recombinases involve the detailed stereochemical interfaces between subunits that regulate the recombination process. While Cre and XerC/D have similar N-terminal domain structures and might be expected to form similar protein-protein contacts in the recombination assembly that could mediate an allosteric regulatory process, it is clear that these contacts have not been conserved in Flp recombinase (11a). A discussion of the stereochemical components of this regulatory system, given the available structural and biochemical data, is presented elsewhere (64).

A closely related question for the future regards the isomerization of the HJ-intermediate in tyrosine recombinase recombination. While the isomerization appears to be quite subtle in the Cre system, requiring only small changes in quaternary structure and no branch migration of the junction, the same need not be true of Flp, Int, and XerCD, where the crossover sequence lengths are in most cases different. In the Flp recombinase 8-bp crossover system, both a limited branch migration model and a simple 4-bp swap model would seem possible (65). In the XerCD 8 bp crossover and λ-Int 7 bp crossover systems, the formation of specific nucleoprotein architectures adjacent to the recombination sites leads to interesting questions regarding the role of branch migration in the recombination pathway. The extent to which these nucleoprotein assemblies might inhibit, promote, or simply tolerate a limited branch migration process during recombination (16) will be an important issue in understanding the more complex tyrosine recombination systems in the future.

ACKNOWLEDGMENTS

I gratefully acknowledge the valuable comments, discussions, and support received from members of the site-specific recombination community during my laboratory's work in the Cre-*loxP* system. I also thank Drs. Feng Guo and Michael Gopaul, who carried out the crystallographic studies in the Cre-*loxP* system described here, and the National Institutes of Health for financial support.

LITERATURE CITED

1. Abremski KE, Hoess RH. 1984. Bacteriophage P1 site-specific recombination. Purification and properties of the Cre recombinase protein. *J. Biol. Chem.* 259: 1509–14
2. Abremski KE, Hoess RH. 1992. Evidence for a second conserved arginine residue in the integrase family of recombination proteins. *Protein Eng.* 5:87–91
3. Abremski KE, Hoess RH, Sternberg N. 1983. Studies on the properties of P1 site-specific recombination: evidence for topologically unlinked products following recombination. *Cell* 32:1301–11
4. Arciszewska L, Grainge I, Sherratt D. 1995. Effects of Holliday junction position on Xer-mediated recombination in vitro. *EMBO J.* 14:2651–60
5. Arciszewska LK, Sherratt DJ. 1995. Xer site-specific recombination in vitro. *EMBO J.* 14:2112–20
6. Argos P, Landy A, Abremski K, Egan JB, Haggard LE, et al. 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. *EMBO J.* 5:433–40
7. Blakely G, May G, McCulloch R, Arciszewska LK, Burke M, et al. 1993. Two related recombinases are required for site-specific recombination at dif and cer in *E. coli* K12. *Cell* 75:351–61
8. Burgin AB, Nash HA. 1995. Suicide substrates reveal properties of the homology-dependent steps during integrative recombination of bacteriophage lambda. *Curr. Biol.* 5:1312–21
9. Campbell A. 1962. Episomes. *Adv. Genet.* 11:101–45
10. Cao YH, Hayes F. 1999. A newly identified, essential catalytic residue in a critical secondary structure element in the integrase family of site-specific recombinases is conserved in a similar element in eucaryotic type IB topoisomerases. *J. Mol. Biol.* 289:517–27
11. Carson M. 1991. Ribbons 2.0. *J. Appl. Crystallogr.* 24:958–61

11a. Chen Y, Narendra U, Iype LE, Cox MM, Rice PA. 2000. Crystal structure of a Flp recombinase-Holliday junction complex: assembly of an active oligomer by helix swapping. *Mol. Cell.* 6:885–97

12. Cheng CH, Kussie P, Pavletich N, Shuman S. 1998. Conservation of structure and mechanism between eukaryotic topoisomerase I and site-specific recombinases. *Cell* 92:841–50
13. Craig NL. 1988. The mechanism of conservative site-specific recombination. *Annu. Rev. Genet.* 22:77–105
14. Duckett DR, Murchie AIH, Diekmann S, von Kitzing E, Kemper B, et al. 1988. The structure of the Holliday junction, and its resolution. *Cell* 55:79–89
15. Esposito D, Scocca JJ. 1997. The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res.* 25:3605–14
16. Franz B, Landy A. 1995. The Holliday junction intermediates of lambda integrative and excisive recombination respond differently to the bending proteins integration host factor and excisionase. *EMBO J.* 14:397–406
17. Gopaul DN, Guo F, Van Duyne GD. 1998. Structure of the Holliday junction intermediate in Cre-loxP site-specific recombination. *EMBO J.* 17:4175–87
18. Gopaul DN, Van Duyne GD. 1999. Structure and mechanism in site-specific recombination. *Curr. Opin. Struct. Biol.* 9:14–20
19. Grainge I, Jayaram M. 1999. The integrase family of recombinase: organization and function of the active site. *Mol. Microbiol.* 33:449–56

20. Grindley NDF. 1997. Site-specific recombination: synapsis and strand exchange revealed. *Curr. Biol.* 7:R608–12
21. Guo F. 2000. *Structural studies on the mechanism of Cre-loxP site-specific recombination.* PhD thesis. Philadelphia: Univ. Penn. 273 pp.
22. Guo F, Gopaul DN, Van Duyne GD. 1999. Asymmetric DNA bending in the Cre-loxP site-specific recombination synapse. *Proc. Natl. Acad. Sci. USA* 96:7143–48
23. Guo F, Gopaul DN, Van Duyne GD. 1997. Structure of Cre recombinase complexed with DNA in a site-specific recombination synapse. *Nature* 389:40–46
24. Hallet B, Sherratt DJ. 1997. Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. *FEMS Microbiol. Rev.* 21:157–78
25. Hargreaves D, Rice DW, Sedelnikova SE, Artymiuk PJ, Lloyd RG, et al. 1998. Crystal structure of E. coli RuvA with bound DNA Holliday junction at 6 A resolution. *Nat. Struct. Biol.* 5:441–46
26. Hayes F, Sherratt DJ. 1997. Recombinase binding specificity at the chromosome dimer resolution site dif of *Escherichia coli*. *J. Mol. Biol.* 266:525–37
27. Hickman AB, Waninger S, Scocca JJ, Dyda F. 1997. Molecular organization in site-specific recombination: the catalytic domain of bacteriophage HP1 integrase at 2.7 A resolution. *Cell* 89:227–37
28. Hoess RH, Abremski KE. 1985. Mechanism of strand cleavage and exchange in the Cre-lox site-specific recombination system. *J. Mol. Biol.* 181:351–62
29. Hoess RH, Abremski KE, Irwin S, Kendall M, Mack A. 1990. DNA specificity of the Cre recombinase resides in the 25 kDa carboxyl domain of the protein. *J. Mol. Biol.* 216:873–82
30. Hoess RH, Wierzbicki A, Abremski KE. 1990. Synapsis in the Cre-lox site-specific recombination system. In *Structure and Methods*, ed. RH Sarma, MH Sarma, pp. 203–213. New York: Adenine
31. Kitts PA, Nash HA. 1988. Bacteriophage lambda site-specific recombination proceeds with a defined order of strand exchanges. *J. Mol. Biol.* 204:95–107
32. Krogh BO, Shuman S. 2000. Catalytic mechanism of DNA topoisomerase IB. *Mol. Cell* 5:1035–41
33. Kwon HJ, Tirumalai R, Landy A, Ellenberger T. 1997. Flexibility in DNA recombination: structure of the lambda integrase catalytic core. *Science* 276:126–31
34. Landy A. 1989. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem.* 58: 913–49
35. Landy A. 1993. Mechanistic and structural complexity in the site-specific recombination pathways of Int and FLP. *Curr. Opin. Genet. Dev.* 3:699–707
36. Landy A. 1999. Coming or going it's another pretty picture for the lambda-Int family album. *Proc. Natl. Acad. Sci. USA* 96:7122–24
37. Lee J, Jayaram M. 1995. Role of partner homology in DNA recombination. *J. Biol. Chem.* 270:4042–52
38. Lee J, Tonozuka T, Jayaram M. 1997. Mechanism of active site exclusion in a site-specific recombinase: role of the DNA substrate in conferring half-of-the-sites activity. *Genes Dev.* 11:3061–71
39. Lee J, Tribble G, Jayaram M. 2000. Resolution of tethered antiparallel and parallel holliday junctions by the Flp site-specific recombinase. *J. Mol. Biol.* 296:403–19
40. Lilley DMJ. 1997. Site-specific recombination caught in the act. *Chem. Biol.* 4:717–20
41. Lilley DMJ, Clegg RM. 1993. The structure of the four-way junction in DNA. *Annu. Rev. Biophys. Biomol. Struct.* 22:299–328
42. Luetke KH, Sadowski PD. 1995. The role of DNA bending in Flp-mediated

site-specific recombination. *J. Mol. Biol.* 251:493–506
43. Luetke KH, Sadowski PD. 1998. Determinants of the position of a Flp-induced DNA bend. *Nucleic Acids Res.* 26:1401–7
44. Nash HA. 1996. Site-specific recombination: integration, excision, resolution, and inversion of defined DNA segments.In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*, ed. FC Neidhardt, R Curtiss, JL Ingraham, EC Lin, KB Low, et al, pp. 2363–76. Washington, DC.: ASM Press
45. Nunes-Düby SE, Azaro MA, Landy A. 1995. Swapping DNA strands and sensing homology without branch migration in lambda site-specific recombination. *Curr. Biol.* 5:139–48
46. Nunes-Düby SE, Matsumoto L, Landy A. 1987. Site-specific recombination intermediates trapped with suicide substrates. *Cell* 50:779–88
47. Nunes-Düby SE, Tirumalai RS, Kwon HJ, Ellenberger T, Landy A. 1998. Similarities and differences among 105 members of the Int family of site-specific recombinases. *Nucleic Acids Res.* 26:391–406
48. Pan H, Clary D, Sadowski PD. 1991. Identification of the DNA-binding domain of the FLP recombinase. *J. Biol. Chem.* 266:11347–54
49. Redinbo MR, Stewart L, Kuhn P, Champoux JJ, Hol WG. 1998. Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science* 279:1504–13
50. Deleted in proof.
51. Ringrose L, Lounnas V, Ehrlich L, Buchholz F, Wade R, et al. 1998. Comparative kinetic analysis of FLP and Cre recombinases: mathematical models for DNA binding and recombination. *J. Mol. Biol.* 284:363–84
52. Roe SM, Barlow T, Brown T, Oram M, Keeley A, et al. 1998. Crystal structure of an octameric RuvA-Holliday junction complex. *Mol. Cell* 2:361–72
53. Sadowski PD. 1995. The Flp recombinase of the 2-microns plasmid of Saccharomyces cerevisiae. *Prog. Nucleic Acids Res. Mol. Biol.* 51:53–91
54. Sauer B. 1993. Manipulation of transgenes by site-specific recombination: use of Cre recombinase. *Methods Enzymol.* 225:890–900
55. Schwartz CJ, Sadowski PD. 1989. FLP recombinase of the 2 microns circle plasmid of *Saccharomyces cerevisiae* bends its DNA target. Isolation of FLP mutants defective in DNA bending. *J. Mol. Biol.* 205:647–58
56. Schwartz CJ, Sadowski PD. 1990. FLP protein of 2 mu circle plasmid of yeast induces multiple bends in the FLP recognition target site. *J. Mol. Biol.* 216:289–98
57. Sherratt DJ, Arciszewska LK, Blakely G, Colloms S, Grant K, et al. 1995. Site-specific recombination and circular chromosome segregation. *Philos. Trans. R. Soc. London Ser. B* 347:37–42
58. Sherratt DJ, Wigley DB. 1998. Conserved themes but novel activities in recombinases and topoisomerases. *Cell* 93:149–52
59. Stark WM, Boocock MR, Sherratt DJ. 1992. Catalysis by site-specific recombinases. *Trends Genet.* 8:432–39
60. Stark WM, Sherratt DJ, Boocock MR. 1989. Site-specific recombination by Tn3 resolvase: topological changes in the forward and reverse reactions. *Cell* 58:779–90
61. Sternberg N, Hamilton D, Austin S, Yarmolinsky M, Hoess RH. 1981. Site-specific recombination and its role in the life cycle of bacteriophage P1. *Cold Spring Harbor Symp. Quant. Biol.* 1:297–309
62. Subramanya HS, Arciszewska LK, Baker RA, Bird LE, Sherratt DJ, et al. 1997. Crystal structure of the site-specific recombinase, XerD. *EMBO J.* 16:5178–87
63. Tirumalai RS, Kwon HJ, Cardente EH, Ellenberger T, Landy A. 1998. Recognition of core-type DNA sites by lambda integrase. *J. Mol. Biol.* 279:513–27
64. Van Duyne GD. 2001. A structural view of

tyrosine recombinase site-specific recombination. In *Mobile DNA II*, ed. N Craig, R Craige, M Gellert, A Lambowitz. Washington DC: ASM Press

65. Voziyanov Y, Pathania S, Jayaram M. 1999. A general model for site-specific recombination by the integrase family recombinases. *Nucleic Acids Res.* 27:930–41
66. Weisberg RA, Enquist LW, Foeller C, Landy A. 1983. A role for DNA homology in site-specific recombination: the isolation and characterization of a site affinity mutant of coliphage lambda. *J. Mol. Biol.* 170:319–42
67. Wittschieben J, Shuman S. 1997. Mechanism of DNA transesterification by vaccinia topoisomerase: catalytic contributions of essential residues Arg-130, Gly-132, Tyr-136 and Lys-167. *Nucleic Acids Res.* 25:3001–8
68. Yang W, Mizuuchi K. 1997. Site-specific recombination in plane view. *Structure* 5:1401–6
69. Zhu XD, Pan G, Luetke K, Sadowski PD. 1995. Homology requirements for ligation and strand exchange by the FLP recombinase. *J. Biol. Chem.* 270:11646–53

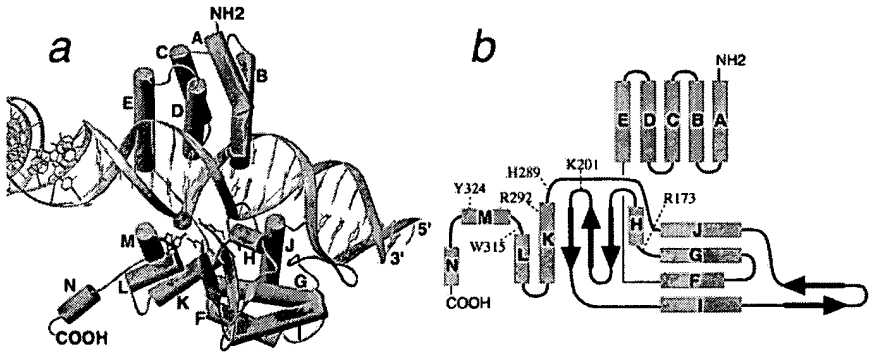


Figure 3 Domain structure and DNA-binding of Cre recombinase. (*a*) Ribbon/cylinder drawing of one Cre subunit bound to a *loxP* half-site (representing one half of Figure 1). Helices are labeled as in (*b*) and active site residues are drawn as red sticks. The scissile phosphate is drawn as a red sphere. (*b*) Folding pattern of Cre recombinase. Helices A through E comprise the N-terminal domain, and the remainder forms the catalytic domain. A similar fold has been observed for the catalytic domains of λ-integrase (33), HP1-integrase (27), and XerD recombinase (62). This fold forms a subset of the eukaryotic topoisomerase IB (12) and the yeast (Flp) recombinase (50) catalytic domains. The locations of conserved catalytic residues (Cre numbering) are indicated. Ribbon/cylinder illustrations in Figures 3–7 were produced with the program RIBBONS (11).

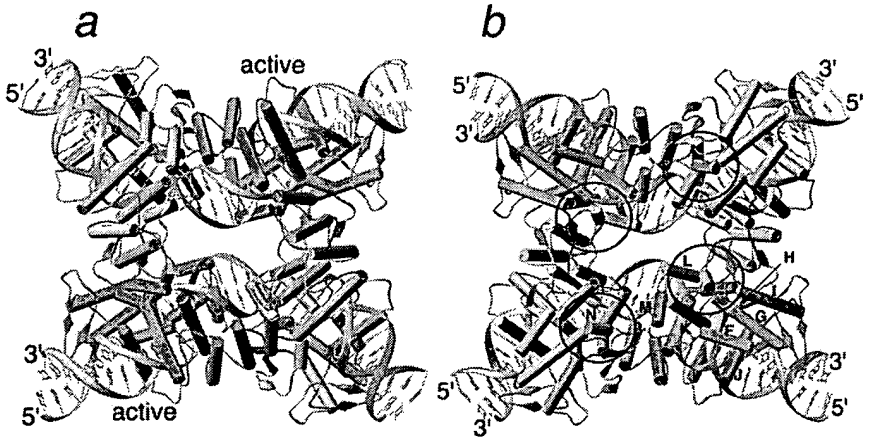


Figure 4 Structure of the synapsed substrates (or synapsed products) Cre-DNA complex (22). (*a*) View of the synapse from the amino-terminal domain side. Helices in the N-terminal domain of one subunit are labeled. Subunits colored green are active for substrate cleavage. (*b*) View of the synapse from the catalytic domain side, rotated 180 degrees about a horizontal axis with respect to (*a*). Helices in one of the active (green) subunits are labeled and the C-terminal helices-N that interact with adjacent subunits are circled. The helix swap is cyclic among the four subunits.

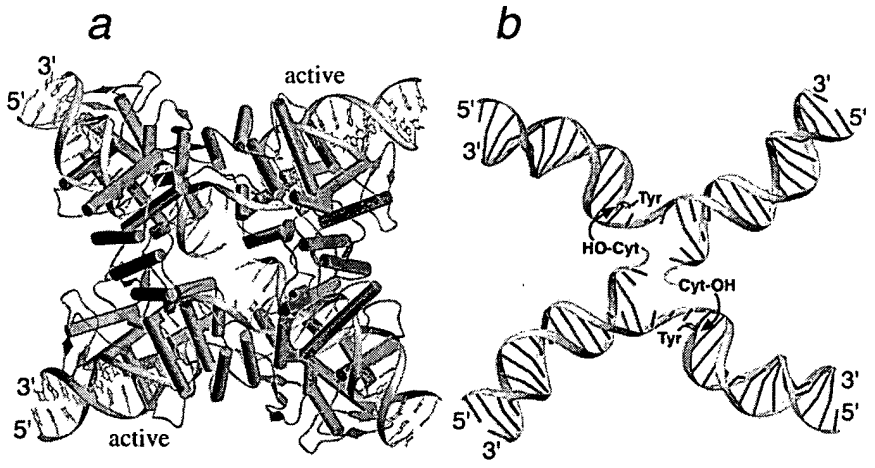


Figure 5 Structure of the Cre-DNA covalent intermediate (23). 5′-cytidine residues in nicked suicide substrates were cleaved and allowed to diffuse away in order to trap this intermediate (46). (*a*) View from the N-terminal domain side, as in Figure 4a. (*b*) The DNA alone in the cleaved intermediate structure, viewed from the catalytic domain side of the complex, as in Figure 4b. Missing 5'-cytidine residues and the direction of strand movement and nucleophilic attack during strand exchange are indicated.

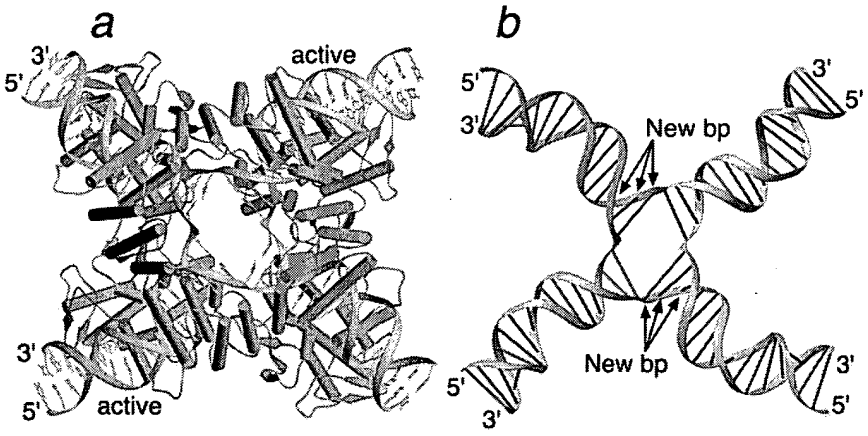


Figure 7 Structure of the Cre-Holliday junction intermediate (17). (*a*) View of the complex from the N-terminal domain side, as in Figures 4a and 5a. (*b*) The DNA in the Cre-HJ intermediate structure viewed from the catalytic domain side as in Figures 4b and 5b. The formation of three base-pairs with the partner substrates as a result of strand exchange is indicated by arrows (compare with Figure 5b). Only the phosphodiester backbone of the exchanged bases needs to move in order to form the HJ intermediate.

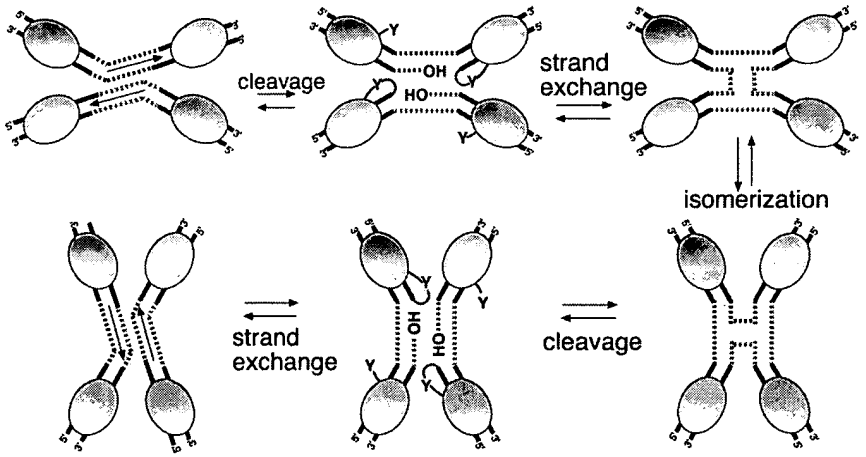


Figure 9 Mechanism of Cre-*loxP* site-specific recombination, based on the strand-swapping model of Nunes-Düby et al (45) and on structural models of Cre recombinase/DNA complexes. Green subunits are active for cleavage in the top half of the pathway, and purple subunits are active for cleavage in the bottom half of the pathway. The DNA substrates lie nearly in the same plane and undergo only a subtle scissoring motion at the HJ isomerization step of the reaction that serves to switch the roles of the protein subunits and switch strands that are activated for exchange. The mechanism does not require branch migration of the HJ intermediate.

Annu. Rev. Biochem. 1989. 58:913–49

DYNAMIC, STRUCTURAL, AND REGULATORY ASPECTS OF λ SITE-SPECIFIC RECOMBINATION

Arthur Landy

Division of Biology and Medicine, Brown University, Providence, Rhode Island 02912

CONTENTS

0066-4154/89/0701-0913$02.00

PERSPECTIVES AND SUMMARY

On the 25th anniversary of Robin Holliday's model for single-strand exchange intermediates in homologous recombination (1), it is especially fitting to review the most recent developments in the site-specific recombination system of *Escherichia coli* bacteriophage λ. This article focuses on developments since the most recent reviews of λ recombination (2, 3), and has a different emphasis than a concurrent review (4) that explores the regulatory aspects of λ recombination in more detail. The present state of the field has its roots in the wealth of elegant λ genetics that rallied around Allan Campbell's model for viral integration (5). The purification of essential proteins by Howard Nash and his collaborators (6–8), and the definition of recombination target sites (9–11), allowed this system to be the first recombination pathway eligible for rigorous biochemical analysis. New results to be discussed in this review concern the dynamic, structural, and regulatory aspects of λ recombination. They are summarized below.

The dynamic aspects of λ recombination (such as mechanisms of strand exchange) prominently feature the biochemical analysis of Holliday junction recombination intermediates. Artificial DNA constructs designed to model putative Holliday-junction recombination intermediates are efficiently and specifically resolved by purified λ Integrase into normal recombinant products (12). Holliday intermediates can also be trapped as the major product in a λ recombination reaction using "suicide recombination substrates" (13). The pair of reciprocal strand exchanges that first form and then resolve the Holliday junctions proceed in a strictly prescribed order that is the same for integrative and excisive recombination (13, 14). This order is determined by DNA sequences and protein binding sites distant from the region of strand exchange (13, 14). Formation and resolution of the Holliday junction proceeds via a high-energy intermediate containing a covalent linkage between the DNA and Tyr-342 of Integrase protein (15). The requirement for a full 7 bp of DNA homology between recombining partners (16, 17) appears not to be for synapsis but rather for the purpose of branch migration between the staggered strand exchange sites within the Holliday intermediate (13, 18) (B. DeMassy, R. A. Weisberg, personal communication). The synapsis of *att* sites is most likely mediated by protein-protein and protein-DNA interactions, and in integrative recombination this may involve a protein-decorated supercoiled "donor" (*att*P) pairing with a protein-free "recipient" (*att*B) (19).

The structural aspects of λ recombination are becoming interesting as models for other complex protein–nucleic acid systems, such as those found in regulation of transcription and DNA replication. The *att* site DNAs function as part of higher-order protein-DNA complexes called "intasomes" (20, 21). These are formed by a set of cooperative and competitive protein-protein

interactions involving four proteins and 15 binding sites. Formation of one of these intasomes (*att*P) requires supercoiled DNA (22). The intasome structure is shaped largely by three sequence-specific "accessory" proteins (IHF, Xis, and FIS) that induce extremely sharp bends in DNA (23). The Integrase protein has the potential to tether distant sequences and form DNA loops by virtue of two autonomous DNA-binding domains with different recognition specificities (24). Integrase executes recombination (7) by a mechanism of strand cleavage and ligation that is shared with more than 15 related recombinases (Int Family) (15, 25).

The extreme directionality and regulation of λ recombination are among its most striking features. Directionality is the consequence of two distinctly different pathways for integrative and excisive recombination. Each pathway uses a unique, but overlapping, set of the 15 protein binding sites that comprise *att* site DNAs (26, 27). Cooperative and competitive interactions involving all four recombination proteins determine the direction of recombination. They also bestow upon the reactions sensitivity to host and viral physiology and to environmental conditions. Regulation of recombination at the level of gene expression is augmented by a mechanism-based response in which host-encoded proteins are incorporated as integral elements of the recombination reaction. One of the host-encoded proteins (IHF) that is required for both integrative and excisive recombination has one binding site that must be occupied for integration and must be vacant for excision (28). The other host-encoded protein (FIS) dramatically stimulates excisive recombination when levels of the phage-encoded Xis protein are low (29). Intracellular levels of FIS, which is also an accessory protein for a different class of recombination reactions, drop by 70-fold as cells go from exponential to stationary phase (29).

In addition to the recent developments outlined above, reference will also be made to FLP and Cre, two well-characterized members of the Int family. Although they are similar to Int in the basic mechanisms of strand exchange, they are quite different in terms of structural complexity, directionality, and regulation.

BACKGROUND

Conservative site-specific recombination is distinguished from homologous recombination and transposition by a high degree of specificity for both partners; the strand exchange mechanism involves the cleavage and rejoining of specific DNA sequences in the absence of DNA synthesis. Two major families comprise this class of reactions. The Resolvase-Invertase family is distinguished by the use of a conserved serine to form covalent intermediates with DNA, a constraint to intramolecular reactions, and an orientation prefer-

ence for recombination sites that can be either direct repeats (e.g. $\gamma\delta$ and Tn-3 Resolvase) or inverted repeats (e.g. the Hin, Gin, and Cin inversion systems).

The other family of reactions (Int Family) (15, 25) is distinguished by the use of a conserved tyrosine to form covalent intermediates with DNA, the ability to execute both inter and intra molecular reactions, and the capacity to recombine sites with either direct or inverted orientation. Members of the Int family can be further subdivided into the minimal systems (e.g. FLP of yeast 2μ circle and Cre of bacteriophage P1) and the complex systems, such as bacteriophage λ and related viruses.

The λ recombination system consists of four recombination sites (*att* sites) and four proteins that carry out the integrative and excisive recombination reactions as shown in Figure 1. The Integrase (Int) protein cuts and reseals DNA to carry out strand exchange. The other phage-encoded protein (Xis) is required only for excisive recombination. Of the two host-encoded proteins, IHF is required for both reactions, while FIS enhances only excisive recombination and only under special circumstances. Integrative recombination between specific sites on the viral (*att*P) and bacterial (*att*B) chromosomes generates prophage bounded by the recombinant products *att*L and *att*R. An excisive recombination between *att*L and *att*R regenerates *att*B and *att*P in the bacterial and excised viral chromosomes. Of the four *att* sites, *att*P (POP') is the most complex and *att*B (BOB') is the simplest (see Figures 1 and 2 for the coordinates and protein binding sites). The prophage sites *att*L (BOP') and *att*R (POB') are "hybrids" of *att*P and *att*B (see Figure 1).

The reader is referred to other closely related, or overlapping reviews on regulation of λ recombination (4), conservative site-specific recombination pathways (30, 31), Resolvases (32, 33), Invertases (34), enhancers (35), resolution of Holliday junctions (36, 37), FLP (38), topology (39–41), and to many related chapters in the books *Transposition* (42), *Genetic Recombination* (43), and *Mobile DNA* (44).

PROTEINS AND THEIR GENES

Int

Int is a basic protein of 40,330 M_r (45, 46) that has type I topoisomerase activity (6, 7, 15, 47, 48) and binds specifically to two different families of DNA sequences (49, 50). During recombination, it carries out the cutting and resealing of *att* site DNA via a covalent Int-DNA intermediate in the absense of any high-energy cofactors (8, 51, 52). Int can also transfer covalently bound DNA to the 5' OH of another DNA molecule in a sequence-independent reaction (S. E. Nunes-Düby, A. Landy, unpublished results). Int is encoded in the λ genome adjacent to the phage *att* site, reflecting the

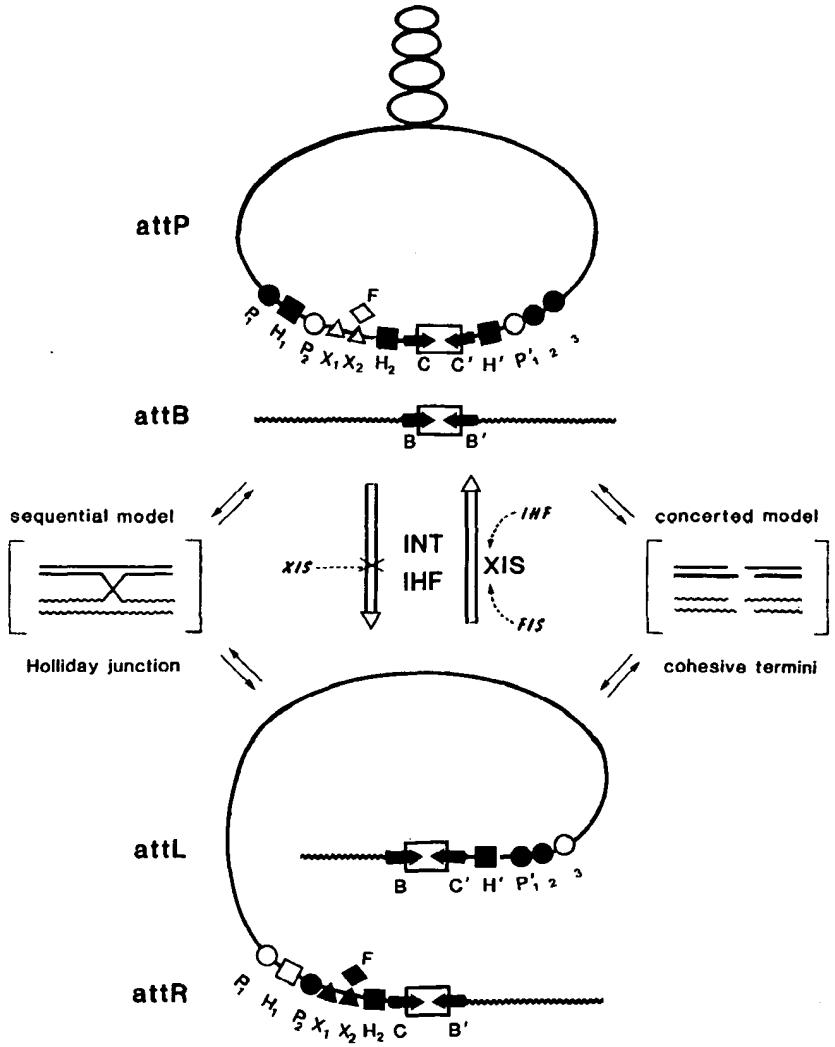


Figure 1 Integrative and excisive recombination pathways. The protein binding sites for arm-type Int (○), core-type Int (▬▶), IHF (□), Xis (△), and FIS (◇) are indicated by filled symbols when that site is occupied by its cognate protein to make a competent recombination partner for integrative (⇓) or excisive (⇑) recombination. Proteins required for each reaction (Int, IHF, and Xis) are in bold, proteins that inhibit (Xis and IHF) or enhance (FIS) the indicated reactions are in italics. Of the two models for strand exchange ([ ]), the sequential model (*left*) is correct.

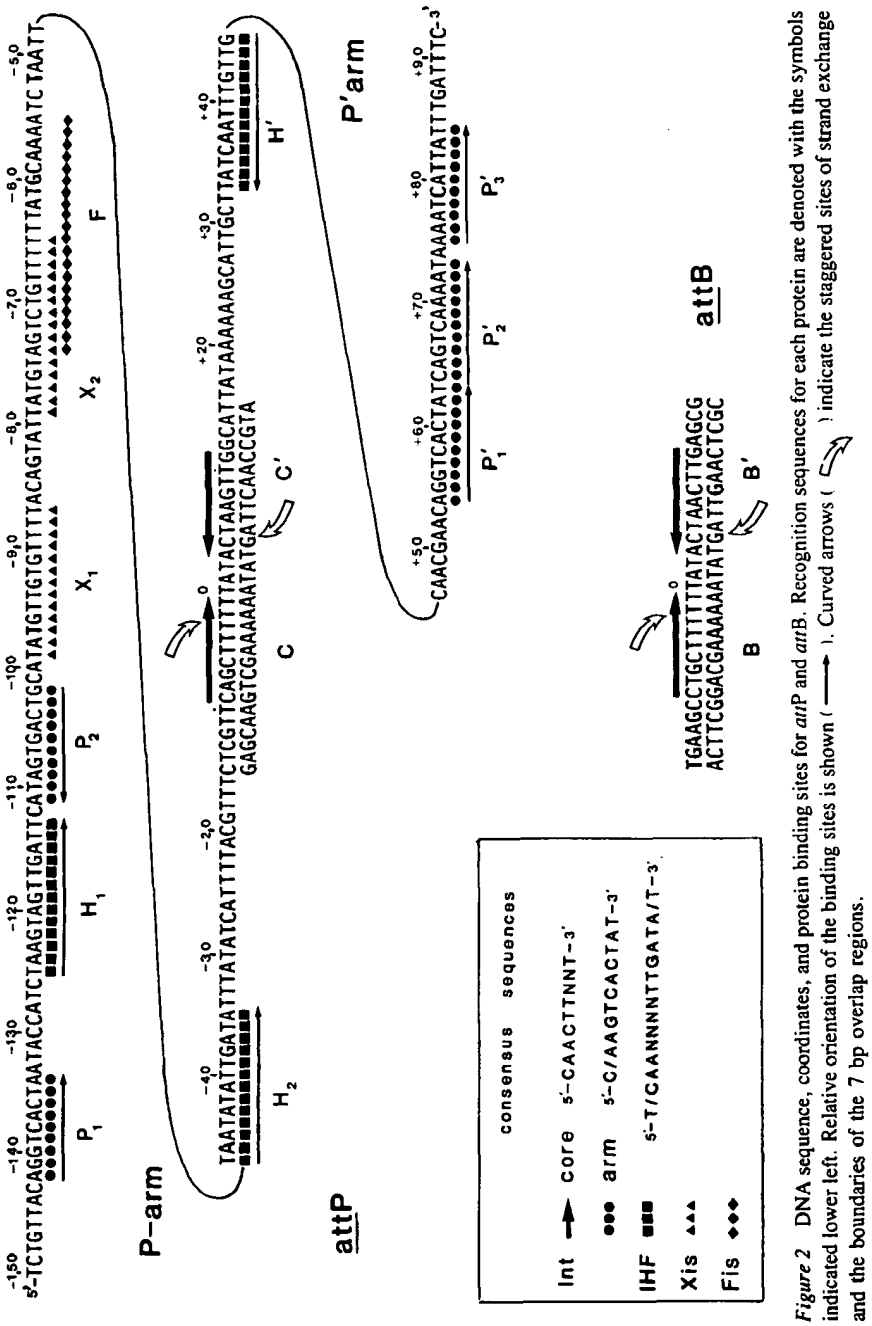


Figure 2 DNA sequence, coordinates, and protein binding sites for *attP* and *attB*. Recognition sequences for each protein are denoted with the symbols indicated lower left. Relative orientation of the binding sites is shown (——→). Curved arrows ( ) indicate the staggered sites of strand exchange and the boundaries of the 7 bp overlap regions.

common motif of recombinase genes that are very close to their sites of action. The specific arrangement of recombination genes often reflects their respective strategies for regulation of gene expression, which, in the case of λ *int*, involves retroregulation (53–57), its own cII-dependent promoter (58), and transcription from the distant P_L promoter [reviewed in (4)]. In phages P22 and P2, the *int* genes are oriented as in λ, but in ø80 and P4 they have the opposite orientation (59–61). The P2 *int* gene, which was once hypothesized to be split by the *att* site (62), has recently been shown to terminate its coding region just before *att* (61).

Thus far only one classical *int* mutant (resulting in a Glu to Lys change at position 174) (63) has been well characterized biochemically. The mutation was isolated in two independent experiments: (*a*) by selection for λ bacteriophage that could undergo site-specific recombination in an *E. coli* host mutant for IHF (called *int*-h) (64); (*b*) by selection for an Int partially independent of Xis in excisive recombination (called *xin*) (C. Gritzmacher, L. W. Enquist, R. A. Weisberg, unpublished data). In the absence of IHF, where Int^+ activity is depressed approximately 500-fold, Int-h activity is only depressed approximately 10-fold (52). The fact that this reduced, but significant, Int-h activity is identical for supercoiled and relaxed substrates indicates that the normal supercoiling requirement for *att*P is determined largely by IHF. The two most plausible explanations for this phenotype are that Int-h/Xin may have a higher affinity than Int^+ for core-type binding sites; or that it works differently on the synaptic intermediate, either to stabilize it or accelerate its conversion to recombinant product (65). A number of other *int* mutants that are interesting because they are recombination proficient (66) or dominant (63) have been studied but not thoroughly characterized.

A comparative study of the primary structures of proteins from seven different bacteriophages pointed to an "Integrase family" of recombination proteins that now includes more than 15 members (15, 25). Within this family of proteins there is one region of approximately 40 amino acids near the carboxy-terminus where at 25 positions more than 50% of the residues belong to the same amino acid exchange group (67). Particularly striking are three perfectly conserved residues within this region: His-308, Arg-311, and Tyr-342 (using the numbering of λ Int).

The biochemical significance of this region was determined using a family of suicide recombination substrates designed to accumulate transient recombination intermediates (13) (see below, SYNAPSIS AND STRAND EXCHANGE, and Figure 3). It was shown that Tyr-342 of Int forms a covalent bond with DNA at the sites of strand exchange. A mutant Int in which Tyr-342 is changed to phenylalanine is devoid of both topoisomerase and recombinase activity but still binds to both classes of Int DNA-binding sites with an affinity comparable to wild-type Int (15). The applicability of these

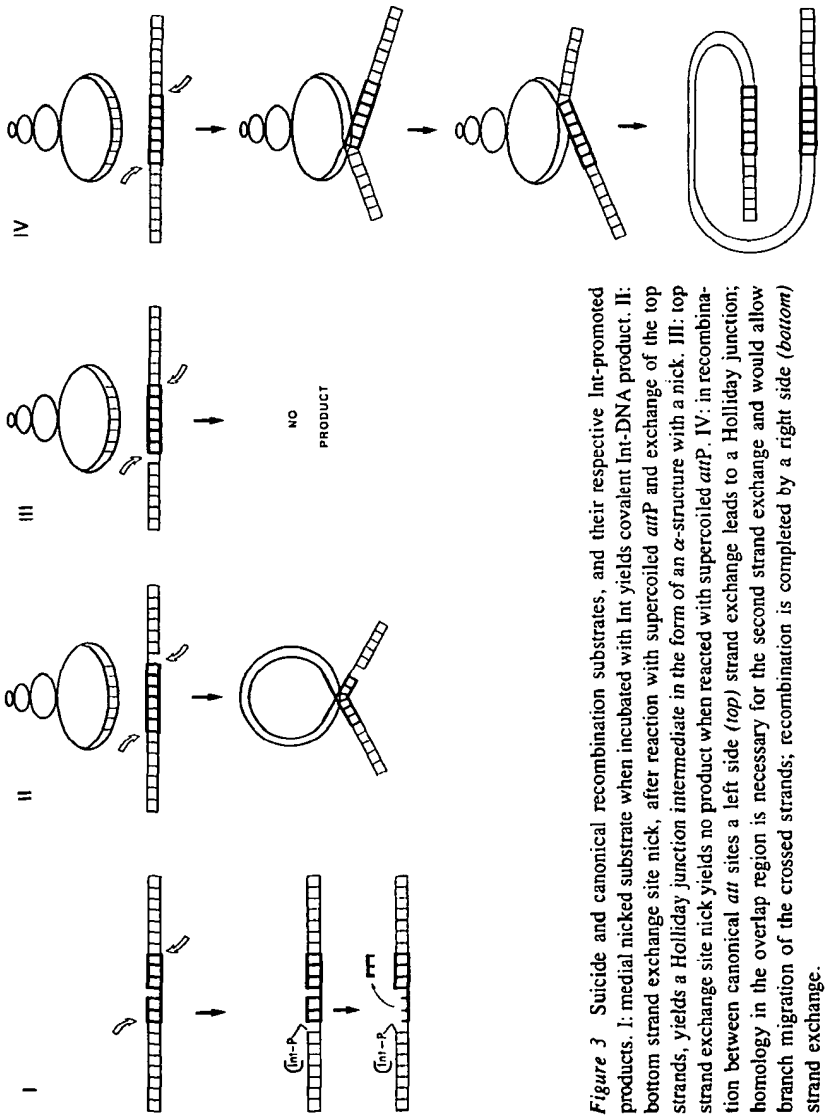


Figure 3 Suicide and canonical recombination substrates, and their respective Int-promoted products. I: medial nicked substrate when incubated with Int yields covalent Int-DNA product. II: bottom strand exchange site nick, after reaction with supercoiled *attP* and exchange of the top strands, yields a Holliday junction intermediate in the form of an α-structure with a nick. III: top strand exchange site nick yields no product when reacted with supercoiled *attP*. IV: in recombination between canonical *att* sites a left side *(top)* strand exchange leads to a Holliday junction; homology in the overlap region is necessary for the second strand exchange and would allow branch migration of the crossed strands; recombination is completed by a right side *(bottom)* strand exchange.

results to the Int family as a whole is supported by studies on FLP and Cre (68–70). In contrast, the Resolvase-Invertase family of recombinases forms a covalent linkage with DNA via a Ser(p) (71) located very close to the amino-terminus (35, 71–73) (see also BACKGROUND).

The recombinases within the Int family are found in a rather wide range of biological functions, such as phage integration and excision, fimbrial phase variation, and plasmid maintenance in *E. coli* and yeast. Some members rely on accessory proteins while others are self-sufficient. Some members, like Cre and FLP, are recombinases with a single DNA recognition site, while others, like λ Int, have two different DNA-binding domains. This intrafamily heterogeneity is reflected by the fact that only a single highly conserved 40-amino-acid region is common to all members of the Int family. It has been proposed that the unifying feature of the Int family is the protein domain for the DNA cutting-ligating activity; the identification of Tyr-342 as the active nucleophile in Int lends strong support to this view (15).

The fact that Int recognizes two classes of DNA-binding sites (49, 50) (see below and Figure 2), along with the biochemical differences in binding to each class (51, 74), led to the proposal for a new kind of DNA-binding protein (50). The existence of two functionally distinct DNA recognition domains in the Integrase protein was demonstrated by the fact that core-type and arm-type DNA sequences do not compete with each other for binding in nuclease protection experiments (24). These results rule out the proposal (75) that all of the Int binding sites are recognized by Int on the basis of similar structural DNA features. They also indicate that an Int monomer is capable of binding to both classes of sites simultaneously. The "bivalency" of Int provides a simple structural explanation for the ability of some *int* mutants to negatively complement (63, 76–79).

The identity and structural autonomy of the two DNA-binding domains was established by proteolytic cleavage of Int and footprinting analysis of the resulting two major peptides (24). A chymotryptic cleavage between Leu-64 and Thr-65 generates a 32-kd carboxy-terminal fragment that binds core-type sites and a 7-kd amino-terminal fragment that binds arm-type sites. The amino-terminal fragment that binds exclusively to arm-type sites is not required for catalytic function, while the carboxy-terminal peptide retains topoisomerase function and resolves synthetic *att* site Holliday junctions (24). This is consistent with earlier observations that Int can efficiently resolve Holliday structures, even in the absence of arm-type binding sites (12).

Xis

Xis is a small (M_r of 8630), basic, phage-encoded protein that is required for excision but not integration (45, 46, 80). Xis has no significant homologies with other proteins, including analogous excisionases from other lambdoid

phage, and nothing is known about its structure. While Xis is relatively thermostable in vitro, it is rapidly degraded intracellularly by an unknown protease, so that 80% of the activity decays within five minutes (80–92). Int, on the other hand, is quite stable intracellularly.

In the presence of Int and IHF, Xis promotes efficient recombination between *att*L and *att*R. With greater than 100 mM salt, Xis is required for this reaction, while at low salt it is dispensable, although still stimulatory (83). Several lines of evidence indicate that the low-salt Xis-independent reaction proceeds by a different pathway than the Xis-stimulated reaction (84). A similar difference in pathways has been proposed for integrative recombination at high and low salt with supercoiled versus relaxed *att*P (85).

Xis has been found to confer enhanced thermostability on Int protein in vitro (80), but there is no evidence to indicate whether this reflects an interaction relevant to recombination. No catalytic activity has been associated with Xis, and its function is exerted by sequence-specific cooperative binding to two adjacent sites in the P arm (26, 86). As discussed below, binding at these sites introduces a very sharp bend in the DNA and is also associated with cooperative interactions with DNA-bound Int and FIS (23, 29, 87, 88).

IHF

Integration Host Factor (IHF), as the name implies, was discovered as a cellular function essential for λ site-specific recombination, both integration and excision (89–92). IHF is a heterodimer composed of two subunits, with predicted molecular weights of 11,200 and 10,580 for the α and β subunits respectively (8, 93–95). Each of the subunits is very similar in sequence to the type II DNA-binding proteins, a family that includes the major histone-like proteins of *E. coli* (HU) and other bacteria [reviewed in (96)]. The only known (primary) function of IHF is its ability to bind (97–99) and to bend (23, 100–103) DNA at specific sites.

It may be valid to infer some basic features of IHF structure from a comparison with the HU protein from *Bacillus stearothermophilus* (HBs), whose crystal structure has been solved to 3 Å resolution (104). Both of the IHF subunits show significant homology to the subunit of the HBs homodimer (94). The HBs structure has two intertwined subunits that form a hydrophobic wedge-shape body and two long flexible arms extending from the wide portion of the wedge. It is proposed that the two flexible arms form a right-handed helix that could bind within the major or minor groove of double-stranded B DNA.

Craig & Nash (97) have proposed that IHF binds in the minor groove of DNA based on the patterns of protection against modification by dimethylsulfate (DMS). H. A. Nash (personal communication) has found addi-

tional support for this proposal in the patterns of protection against cleavage of the sugar-phosphate backbone by hydroxyl radicals. In these experiments, which were done on the three IHF sites of *att* DNA, as well as in nuclease protection experiments on other IHF sites (98, 99), IHF is found to protect three to four helical turns of DNA. Despite this relatively large protected region, preliminary stoichiometric data indicate that only a single IHF dimer binds per site (H. A. Nash, personal communication). It is also interesting to note that while IHF has high sequence specificity, the closely related HU and HBs proteins have little or no sequence specificity.

The genes for IHF have been mapped (89, 90), cloned (93–95, 105), and overexpressed (106). The α-encoding gene, *him*A, maps to minute 37, and the β-encoding gene, which is referred to as both *hip* and *him*D, maps to minute 20 on the *E. coli* chromosome. *Hip*/*him*D is part of a gene cluster with at least two promoters (E. L. Flamm, R. A. Weisberg, personal communication). The *him*A gene is part of a more complex cluster whose transcription involves at least seven promoters and three terminators, with the regulation being influenced by growth rate and the SOS response. For a recent review of the regulation see (4).

Some evidence suggests that levels of IHF may vary as a function of cellular physiology, but this point is not clear. A large increase in IHF levels is detected in stationary-phase cells, but this depends on the conditions used to extract the protein (27, 28). Attempts to resolve this question by in vivo chemical modification have not been definitive because all three of the IHF sites in *att*P that were monitored were fully occupied under the growth conditions tested (29).

The lambdoid phages ø80, P22, and λ all have IHF-binding sites within their very different *att* site DNA sequences (98). ø80, like λ, is unable to lysogenize *him*A or *hip*/*him*D mutants (107), while P22 has not been tested. Since the number of IHF-binding sites is different in each of the phage *att* sites, it will be interesting to compare the three recombination reactions at a mechanistic level. (Cre and FLP are two members of the Int family that do not require IHF.) Transposons are another class of recombination elements where a role for IHF has been demonstrated or implicated (99, 101, 108, 109). In some recombination systems an IHF requirement or HU requirement can be replaced with low efficiency by the other protein (108, 110). The roles of IHF extend well beyond recombination and are discussed thoroughly in a recent review (111).

FIS

A second host-encoded protein (in addition to IHF) involved in the λ site-specific recombination reaction was first discovered by testing *E. coli* extracts for proteins that would bind to *att* site DNA in a gel mobility shift assay. Such

an activity was found and subsequently shown to be FIS (29), a protein that had been identified (112–114), and purified (115, 116), on the basis of its ability to stimulate three recombination inversion reactions from the Resolvase-Invertase family. In the λ system, FIS stimulates in vitro excision approximately 20-fold when Xis levels are limiting, however it cannot replace Xis (29). FIS has no effect when Xis is at saturating levels and it does not influence integrative recombination. The in vivo relevance of FIS has been established by demonstrating occupancy of its binding site in the P arm (29), as well as dramatic differences in excision efficiency for FIS^+ and FIS^- cells (R. Johnson, personal communication). When bound to its single binding site in the P arm, FIS induces a DNA bend of approximately 90° (23).

FIS, like IHF, is a small, basic, heat-stable DNA-binding protein (115, 116). Unlike IHF, which is a heterodimer, FIS is a homodimer in solution. The protomer, based on its gene sequence, has 98 amino acids and a calculated pI of 9.5 (117, 118). There is no sequence homology between FIS and the type II DNA-binding proteins such as HU or IHF, despite their gross similarity in physical properties and somewhat analogous roles as bending and accessory proteins in several different recombination pathways. FIS has no cysteines or histidines so that a "zinc-finger" mode of DNA recognition is unlikely (117). However, within the carboxy-terminal portion, at amino acids 74–93, there is a sequence with a high probability of adopting the helix-turn-helix DNA binding motif (117, 118) that has been characterized for repressors and other proteins (119).

The Hin system of *Salmonella typhimurium*, the Gin system of bacteriophage Mu, and the Cin system of bacteriophage P1 constitute a closely related family of FIS-dependent site-specific recombination systems. These systems specifically invert a segment of DNA and thereby mediate expression of two alternate forms of a flagellar antigen or phage tail fiber protein (34, 35, 120). Their requirement for the bacterial protein FIS is mediated by a binding site on the DNA that has been called a "recombinational enhancer." The function of recombinational enhancers, which contain two or three FIS binding sites, is relatively independent of position or orientation relative to the sites of recombination (112–114, 118, 121, 122).

The *fis* gene has been mapped to approximately 72.5 min on the *E. coli* chromosome, between *fab*E and *aro*E (117, 118). The fact that the initiating methionine is preceded by a poor Shine-Delgarno translation initiation site, coupled with the high number of rare codons (11%), may contribute to the low abundance of FIS in *E. coli*. A null mutation in *fis* leads to a 10^3–10^4-fold reduction of Hin and Gin inversion in vivo and a complete loss of detectable FIS activity when extracts are assayed for ability to stimulate inversion in vitro (117, 118). These null mutants also exhibit a 10^3–10^4-fold reduction in λ excisive recombination efficiency, suggesting that under these experimental

conditions, intracellular levels of Xis are limiting (R. Johnson, personal communication). Strains carrying these same null mutations show no reduction in viability, even when this mutation is combined with a mutation in *hip/him*D (encoding one of the subunits of IHF). The inability to find major effects of *fis* mutations on cell physiology, or on the expression of other cellular genes, may reflect the short time these mutants have been available for study.

One reason for predicting some interesting effects of FIS on cell physiology, and a possible role in global aspects of regulation, is that its intracellular concentration drops more than 70-fold as *E. coli* goes from exponential growth to stationary phase (29). It is not known what controls the degradation of FIS or whether gene expression is also involved.

ATT SITE STRUCTURE AND PROTEIN BINDING SITES

The most prominent feature of the four *att* site DNA sequences is a 15 bp "core sequence" they have in common. Individual sequence elements of functional significance lie within, outside, and across the boundaries of this common core, which itself has no functional significance. Controlled resection experiments established that *att*P extends from −150 to +85 from the center of the core, while *att*B extends from approximately −15 to +15 (see Figure 2) (10, 11, 123). The prophage sites generated by integrative recombination, *att*L and *att*R, are fully competent for excisive recombination, but they contain more than the minimal sequences necessary for excision (27) (see below, DIRECTIONALITY AND REGULATION). It is interesting to note that the *att*B sites of six different systems are located within the 3'-terminal regions of various tRNA genes (60, 98) (W. Reiter, P. Palm, S. Yeats, personal communication). While this association with tRNA genes does not hold for the majority of *att*B sites, it is significant and may offer some useful insights about evolution and/or mechanisms.

The *att*P sequence is 73% A+T, with one 48 bp region being 90% A+T (9, 10). As might be expected from this base composition, there are several runs of up to six consecutive adenines, a sequence feature known to generate intrinsic curvature within the DNA (124–126). The center of *att*P curvature has been mapped to −45 to −50 in the P arm (127). This same region is also hypersensitive to modification by bromoacetaldehyde, but only if the DNA is supercoiled (128). While the intrinsic curvature in this region of the P arm may enhance the efficiency of recombination, it is not absolutely required (127). Its function may be to position the *att* site to an external, easily accessible loop in plectonemically supercoiled DNA (129, 130).

In vitro integrative recombination reactions traced the observed requirement for negative supercoiling to the *att*P partner (131). Below 50 mM NaCl

the reaction does not require supercoiled *att*P (132); however, several lines of evidence suggest that this low-salt reaction with linear *att*P proceeds by a different recombination pathway (85). In contrast to integration, excision does not require supercoiling (133), although supercoiling of either *att*L or *att*R stimulates the reaction at limiting protein concentrations. Excision is also less sensitive than integration to salt concentration (132).

Overlap Region

The 7 bp between the staggered cuts responsible for strand exchange is referred to as the "overlap region" (3), since in recombinant molecules this region receives one DNA strand from each partner. The *att* site mutants of Shulman & Gottesman (134) that segregate to both recombinant products have been located within the overlap region by DNA sequencing (135). These mutants greatly impair recombination efficiency by deleting one bp from the overlap region. Insertions of one bp are not as detrimental (136), but a systematic test of spacing has not been carried out. Because this region receives one DNA strand from each partner during recombination, one would predict that it must have the same sequence in both partners. Indeed, many sequence changes can be accommodated within the overlap region as long as the same changes are introduced into both partners (17, 81). The role of DNA:DNA homology will be discussed below (SYNAPSIS AND STRAND EXCHANGE).

An overlap region, defined both by the position of staggered cuts and by the requirement for DNA:DNA homology, is also found in the *lox* sites (Cre system) and FRT sites (FLP system), 6 bp in the former (137–141) and 8 bp in the latter (142–145). In both of these systems some specific sequences within the overlap region have been observed to reduce recombination despite the preservation of homology between the partners. The weaker effect of specific overlap region sequences in λ *att* sites may be due to the dominating effects of the proteins bound to the P and P' arms. Similarly, in the absence of the asymmetry conferred by the P and P' arms of λ, the orientation of *lox* and FRT recombination partners relative to one another is determined wholly by the overlap region sequence (139, 143, 145, 146). The Resolvase-Invertase systems also make use of staggered cuts during strand exchange, but the size of the overlap region is only 2 bp (71, 73, 112, 147, 148).

Int Binding Sites

There are seven Int-binding sites in *att*P, two of which define the outermost limits of *att*P (10, 49). Sequence inspection suggested that these sites could be grouped into two distinct families of sequences: "core-type" (also called "junction-type") sites that are adjacent to the points of strand exchange and "arm-type" sites that are distal to the region of strand exchange. This distinc-

tion was confirmed, and reliable consensus sequences were obtained for each family, by analyzing a number of fortuitous Int binding sites in non-*att* DNA (49, 50). As described above, each family is recognized by one of two autonomous DNA-binding domains of Int (24).

CORE-TYPE SITES The core-type sites (C and C' in *att*P and B and B' in *att*B) are found as inverted repeats at the two ends of the overlap region and include the phosphodiester bonds that are cut by Int during strand exchange. The two base pairs flanking this phosphodiester bond are not specified in the consensus recognition sequence of seven bp (50). This lack of specificity may be a consequence of protein structural features in the catalytic region of Int. Int bound at the core-type sites protects DNA against attack by dimethyl sulfate in both the major and minor grooves along one face of the DNA helix; the inverted repeats face each other across the central major groove. The opposite face of the DNA helix appears to be devoid of strong protein contacts (50). The appropriate configuration of core-type sites is the only requirement for the resolution of synthetic Holliday junctions by Int, or its carboxy-terminal domain (12) (C. Pargellis, B. Franz, and A. Landy, unpublished results).

Although affinity constants for Int binding have not been measured, they are considerably lower for the core-type sites than for the arm-type sites, as estimated by their respective behavior in nuclease protection (19, 49) and gel mobility shift (23) experiments. The relative order of binding affinity for core-type sites is C' > C = B > B' (50). However, the differential in affinities does not appear to be essential, since different sites can be substituted for one another with little or no effect on recombination (13, 26). The differences in affinities of the individual sites may be partially masked by the significant amount of cooperativity between Int proteins bound at two core-type sites (50).

The amount of Int-induced DNA bending at the COC' core-type sites has been estimated to be approximately 17° (23). However, this value must be viewed cautiously because of the relatively low affinity of Int for these sites (50). Additionally, measurements on the isolated core-type sites may not be relevant to the functional recombinogenic complexes (see below, COOPERATIVE INTERACTIONS).

Two core-type sites flanking the overlap region comprise *att*B. This configuration is very similar to the minimal recombination sites of both partners in the Cre and FLP systems. The simplicity of *att*B explains the existence of secondary *att* sites in the *E. coli* chromosome. These sites are utilized for λ integration (at greatly reduced efficiencies) when *att*B is deleted (149–151). Sequence comparisons of the secondary *att* sites indicates that they are degenerate facsimiles of the simple *att*B site. The efficiency of a particular

secondary *att* site is presumably determined by the extent to which it emulates the canonical spacing and sequence of the overlap region and core-type Int binding sites (3, 50). Similarly, fortuitous facsimiles of *att*B, or core-type sites, could comprise the targets for Int topoisomerase activity on supercoiled DNA not containing an *att* site.

ARM-TYPE SITES The initial characterization of a distinct class of essential Int-binding sites well removed from the region of strand exchange was both intriguing and challenging in terms of recombinogenic structures and mechanisms (49, 74). Evidence for how they function in site-specific recombination is beginning to emerge. The five arm-type Int-binding sites have a consensus recognition sequence of seven contiguous base pairs (49). The approximate binding affinities for individual sites indicates a hierarchy of P'1 = P1 > P2 (49, 87, 88). P'2 and P'3 have not been examined in isolation and all three of the P' sites are subject to cooperative binding effects amongst themselves (L. Moitoso de Vargas, A. Landy, unpublished results). In addition to these cooperative effects, sequences outside of the binding site can also influence the binding affinities. [However, the low affinity of Int for P2, which has the best match to the consensus sequence, is not due to context effects (87).]

Classical genetic techniques have produced only one protein-binding mutant in the *att* sites. This mutation, *hen*, is in the P'3 site and changes a completely conserved C to T (152). It is especially interesting because it abolishes integrative recombination without affecting excisive recombination (152). To examine the roles of other arm-type sites, the single base transition of the *hen* mutation has been introduced into each of the five arm-type sites (153). Multi-site changes have also been introduced into a subset of the arm-type sites (27, 88). From these studies and resection experiments (10, 11), it is clear that the P1 and P'2, and P'3 sites are required for integrative recombination, while the P2 site is not. The apparent dispensability of P'1 for integrative recombination is subject to the same reservation that applies to all point mutants without a phenotype in highly cooperative systems.

In excisive recombination, none of the point mutations has as large an effect as seen with integrative recombination (153). The largest effect is observed with the P'1 mutation, suggesting that this site is required for excision, with some of the defect being compensated by cooperative interactions with other sites. P1 is clearly not required for excision, because when it is deleted the *att*R works as well or better than wild-type *att*R (27). While P2 is also dispensable under some circumstances, it does appear to be important for excision. In the absence of the P2 site, more Xis is required to obtain the same level of recombination as with a wild-type *att*R (27, 87). Additionally, the reaction between an *att*R-*hen* P2 and an *att*L-*hen* P'2 or

*att*L-*hen* P'3 is reduced relative to a wild-type *att*L (153). Both of these observations are reflective of cooperative interactions to be discussed below. See Figure 1 for a summary of the site requirements for integrative and excisive recombination.

Xis Sites

Xis protein protects a single region of approximately 40 bp in the P arm (86). Gel mobility shift assays have shown that highly cooperative binding between at least two Xis molecules takes place (26). Mutagenesis of the binding region showed that each half could bind Xis independently, although with lower affinity than the canonical site (26). Protection patterns against DMS modification also provide evidence for two binding sites, denoted X1 and X2, and further suggest that they are arranged as a direct repeat (86). Determining a consensus recognition sequence will require the analysis of fortuitous Xis recognition sites and/or a mutational study of the canonical sites.

The Xis mutations isolated thus far include a single base deletion between the two sites and a series of resected *att* sites in which varying extents of the P arm have been replaced with heterologous DNA. In the first case, excisive recombination is abolished. In the second, it was shown that both X1 and X2 must be occupied; defects in X1 are only overcome when sufficient Xis is present to bind to the mutated as well as the normal site (27).

Xis binding to X1 or X2 bends the DNA approximately 45° and 95°, respectively. When X1 and X2 are both occupied, as is normally the case, the bending angle is greater than 140° (23). This very sharp bend is undoubtedly a key feature in the role of Xis during recombination.

IHF Sites

There are three IHF sites in *att*P: H1 and H2 in the P arm and H' in the P' arm (97). IHF binding to multiple sites on the same DNA fragment is not cooperative (28, 88, 97, 101, 154). Derivation and refinement of a consensus recognition sequence has been aided by the analysis of IHF binding sites from a variety of phage, plasmid, and bacterial sources (97–100, 154–156). IHF protection patterns against modification by dimethyl sulfate further delimit those positions involved in binding (97, 100) (see Figure 2).

As discussed above, and similar to Xis, the only known function of IHF is its ability to bend DNA (23, 100–103). The extent of IHF-induced bending is estimated to be greater than 140° at each of the three *att*P binding sites (23). The strong IHF-induced bend is consistent with the observations and analyses of DNA bending induced by CAP protein (157). The large effects (greater than 100-fold) on IHF binding affinity by sequence changes outside the consensus region (101) (L. Moitoso de Vargas, A. Landy, unpublished results) could be the result of changes in the bendability, and/or intrinsic

curvature, of the DNA. Similarly, variation in the amount of DNA protected by IHF at each of several different binding sites (97, 98) might reflect differences in the length of DNA involved in making the curve around IHF protein. The H2 and H' IHF binding sites, as well as some non-*att* IHF sites (100), occur in regions where the DNA has a substantial intrinsic curvature in the absence of bound protein (23, 127), e.g. as determined by gel mobility shift permutation analyses (125). While intrinsic curvature may have the potential to favor IHF binding if it is in phase with the IHF-induced bend (157), it is not required. A number of IHF sites, such as H1, do not show any evidence of intrinsic curvature (23, 102).

Despite the gross similarity of H1, H2, and H' in the extent of IHF induced bending, the three sites are not functionally equivalent in recombination. Whereas H1 is required for integrative but not excisive recombination, H2 and H' are required for both reactions (26–28, 154). Additionally, it appears that the IHF requirement at H2 and H', but not at H1, can be partially substituted by another protein, possibly HU (154). Progress toward answering some of the questions about IHF should come from the recent IHF-DNA co-crystals made with synthetic oligomer (K. Appelt, personal communication).

FIS Sites

FIS binds to a single site, called F, that overlaps the X2 site in the P arm; the region protected against nuclease digestion is slightly larger than X2 (29). In contrast, the recombinational enhancers of the Hin and Gin family consist of two independent FIS binding sites separated by approximately 48 bp (113, 121, 122). The FIS site in λ *att* also seems to differ in sequence from the enhancer FIS sites (122). This lack of similarity suggests that the FIS recognition determinants are not fully understood and may involve specific DNA conformations not readily apparent in a simple sequence alignment.

Even though FIS can substitute for Xis binding at X2, the two proteins do not recognize the same features in the DNA: their respective DMS-modification protection patterns and the boundaries of the regions protected against nuclease digestion are different, and FIS does not bind to the X1 site (29, 86). In order to separate the effects of FIS and Xis, the X1-X2 site was replaced with an X1-X1 site in the P arm, thereby eliminating FIS binding while retaining the Xis binding (29, 127). This FIS^- mutant made it possible to show that, in contrast to its stimulation of excision, FIS has no effect on integrative recombination.

The effect of adding FIS to a nuclease protection or gel mobility shift assay is equivalent in many respects to simply adding more Xis. If Xis is limiting,

the addition of FIS leads to occupancy of the X1 and F sites with the same cooperativity seen for the X1 and X2 sites when Xis is increased in the absence of FIS. FIS binding to F is also similar to Xis binding to X2 in the extent of induced DNA bending, approximately 90° in both cases (23). Furthermore, the increase in the total apparent bend to more than 140° is similar for X1-F occupancy and for X1-X2 occupancy. Finally, as discussed below, all of the cooperative interactions in which Xis participates with Int and IHF occur equally well when FIS is bound in combination with Xis.

COOPERATIVE INTERACTIONS AND HIGHER-ORDER STRUCTURES

The first evidence for the wrapping of DNA in a putative higher-order structure was the appearance of a 10-base-pair repeat in the nuclease protection patterns of Int-DNA complexes (10, 49, 158). Such patterns of enhanced cutting suggest that one face of the helix is oriented toward the solvent, while the other face is less accessible to nuclease because it lies along a protein surface or forms the inside of a loop (159–162). Electron microscopy graphically revealed large protein complexes at the phage *att* site (20, 163) that suggested the involvement of approximately 230 bp of *att*P DNA and 4–8 Int monomers (20). The term *intasome* was coined to describe these complexes that were considered to contain only Int and DNA (20). It is now used as a generic term for protein–*att* site DNA complexes that also contain the other proteins involved in recombinogenic structures. While electron microscopy was not successful in revealing the importance of IHF in complex formation, it did show that Xis protein favored the formation of complexes involving *att*R and *att*P, and that many combinations of complexes could pair with one another (20, 21).

The first evidence that the functional recombinogenic complex involves DNA wrapped in a nucleosome-like structure came from the topological studies of Pollock & Nash (132). Intramolecular recombination between two *att* sites (either *att*P × *att*B or *att*L × *att*R) oriented in opposite directions on a circular supercoiled molecule inverts one segment of the circle with respect to the other and produces knotted products (164, 165). For integrative recombination, under conditions that minimized the knotting due to random interwrapping of the superhelical DNA, approximately one half of the recombinant products are simple circles and the rest are knotted. For excisive recombination, all of the recombinants are simple circles. The excess knotting of the integrative recombination products is postulated to reflect the wrapping of DNA in the *att*P complex (132). Subsequent topological analyses of knotted and catenated recombinant products have confirmed this result and

indicate that the *att*P "intasome" is wrapped with a negative sign, which in the nucleosome cores of chromatin produces a left-handed solenoid (130, 166–168).

Several lines of evidence suggest that the role of the supercoiling requirement for integrative recombination is to facilitate formation of the *att*P intasome. Since only *att*P requires supercoiling, it is not likely that any step subsequent to synapsis, such as strand exchange, is responsible for the requirement (169). The singularity of *att*P is evidenced by the fact that supercoiling is not required for excisive recombination, although it stimulates the reaction (170). Any degree of DNA wrapping in the *att*L or *att*R intasomes should also have a negative sign, since negative supercoiling of either one enhances the efficiency of recombination (133) (W. Bushman, J. Thompson, A. Landy, unpublished results).

In a more direct biochemical analysis, it was shown that binding of Int to P1, and to a lesser extent to P'3, is more resistant to challenge by excess salmon sperm DNA when *att*P is negatively supercoiled (22). This effect requires the presence of IHF, and is seen only on *att*P (i.e. not when the arm binding sites are separated as in *att*L and *att*R). The good correlation between the sign and degree of supercoiling needed to promote these effects, and that needed to promote recombination, supports the notion that the primary role of supercoiling is to favor formation of the intasome (22).

The importance of the spatial orientation of the P1 and H1 sites within the intasome is seen when the helical phase of these sites is altered with respect to the remainder of *att*P by making insertions or deletions in a nonessential intervening region (85). Under normal reaction conditions (where supercoiling is required), substrates with changes of integral multiples of the DNA helical repeat recombine, while substrates with nonintegral changes do not. With nonsupercoiled (linear) *att*P, in low salt conditions, both the helical phasing and the P1 and H1 sites are much less important (although the overall efficiency is lower than with supercoiled *att*P) (85).

The specific interactions responsible for the intasome structures have been deduced by correlations of recombination with multiprotein nuclease protection assays, competition binding studies, site-specific mutagenesis of protein-binding sites, and results from suicide recombination substrates. These interactions are summarized below and illustrated schematically in Figure 4.

The pairwise cooperative interactions between Int molecules bound at the core-type sites was demonstrated by the fact that Int binds much better to core-type sites positioned as inverted repeats 7 bp apart (i.e. the canonical configuration) than to isolated core-type sites (50, 135). Similar cooperativity is also inferred from studies with artificial Holliday junctions, but this configuration includes the potential interactions among Int protomers bound at

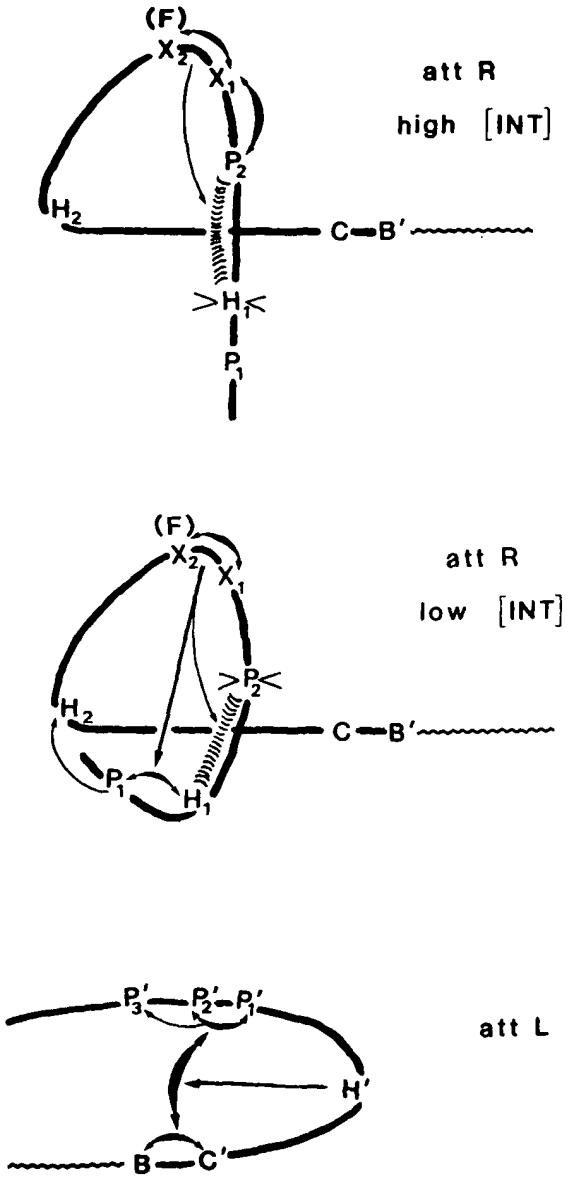


Figure 4 Schematic representation of some of the cooperative interactions observed with linear recombinogenic complexes. In some cases, cooperative interactions ( ) and competitive interactions ( ) are dependent upon binding of another protein ( ). The outcome of a competition at high or low Int is indicated by a vacated ( ) protein binding site. For more information about the individual protein binding sites see Figures 1 and 2.

each of the four core-type sites (B. Franz, A. Landy, unpublished results). Cooperative interactions among Int molecules bound at the core-type sites have been difficult to study because Int has a low affinity for these sites.

The most thoroughly characterized of the pairwise cooperative interactions involves Xis binding at the adjacent X1 and X2 sites (26, 86). This rather strong cooperativity is of special interest for two reasons. First, it involves a head-to-tail arrangement of sites, which is not so common for cooperatively binding proteins. Second, the Xis-Xis cooperativity can be substituted by Xis-FIS cooperativity. Keeping in mind that Xis and FIS do not have similar amino acid sequences, and that they bind to different DNA sequences, it is all the more surprising that for every interaction tested, the X1-F pair works as well as the X1-X2 pair (29). In the following discussion, all Xis interactions refer to Xis bound simultaneously at X1 and X2, and are probably equally applicable to simultaneous binding of Xis at X1 and FIS at F.

Xis bound at X1-X2 (or X1-F binding) cooperatively assists Int binding at P2. In nuclease protection experiments there is a 16-fold reduction in the amount of Int required to fill P2 if X1-X2 is also filled (26). This cooperativity is also observed in recombinations involving *att*R with a defective P2 or X1 site (27, 87, 88).

In addition to the cooperative interactions within the intasome, there are also several competitive interactions. These, however, are not simple pairwise competitions, since any single protein can fill all of its binding sites, and any pairwise combination of proteins can fill all of their respective binding sites (26, 29, 49, 86, 88, 97). Int binding at P2 and IHF binding at the neighboring H1 site are perfectly compatible in the absence of Xis. However, when X1-X2(F) is filled, then binding at H1 is competitive with binding at P2 (29, 88). Clearly, this competition is not due to simple occlusion of a binding site by neighboring proteins, given the pairwise compatibility of protein binding. Rather, the favored interpretation is a competition for adjoining space within a higher-order structure.

In order to better understand the dynamics of the H1-P2 competition, it is necessary to also refer to a complex cooperative interaction involving the P1 site (88). In linear molecules, Int binding to P1 in *att*R is strongly enhanced by the simultaneous occupancy of H1, H2, and X1-X2(F). However, in *att*P this cooperativity is not observed (the P' arm is involved in a different set of cooperative interactions that includes the core-type Int binding sites) (L. Moitoso de Vargas, A. Landy, unpublished). Thus, the binding affinity of Int for P1 in a higher-order complex is determined by interactions occurring on DNA at least 150 bp distant (see Figure 4). Richet et al (22) have found IHF-dependent cooperative interactions involving the P1 site in *att*P that require supercoiling. Thus, two major features that distinguish integrative and excisive recombination (the supercoiling requirement for *att*P and the Xis

requirement for *att*R) are also needed for the long-range interactions involving the P1 site.

The above discussion indicates that each partner in the H1 vs P2 competition is involved in a separate set of cooperative interactions. Therefore, the outcome of the competition is determined by the configuration of *att* site arms, i.e. *att*R vs *att*P, by the relative concentrations of each protein, and by the topological state of the DNA. As will be discussed (DIRECTIONALITY AND REGULATION), the cooperative transition in binding of proteins from the P1, H1, X1/X2(F), H2 set to the P2, X1/X2(F), H2 set (see Figure 5) occurs over the same range of protein concentrations that is required to overcome IHF inhibition of excision and to favor efficient excisive recombination (28, 88).

In contrast to the complexity of the P arm, the relative simplicity of the P' arm highlights one of the primary interactions responsible for intasome structure. In *att*L, IHF bound at H' enhances binding of Int to core-type sites as long as the P' arm-type sites are present. This same cooperative interaction is reflected in an (H'-P')-enhanced nicking of *att*L suicide recombination substrates. In *att*L, both the IHF-enhanced Int binding at core-type sites and recombination proficiency depend upon the correct helical phasing of the P' arm-type sites with respect to the core region (L. Moitoso de Vargas, S. H. Kim, A. Landy, unpublished). These cooperative interactions lead to the architectural element shown in Figure 6 and discussed below.

The number and complexity of the intasome-forming interactions stands in sharp contrast to the compact "minimized" reactions of some other Int family members such as Cre and FLP. Since these are highly efficient site-specific recombinases, it is appealing to ascribe much of the additional complexity of the λ system to the requirements for directionality and regulation, as discussed below.

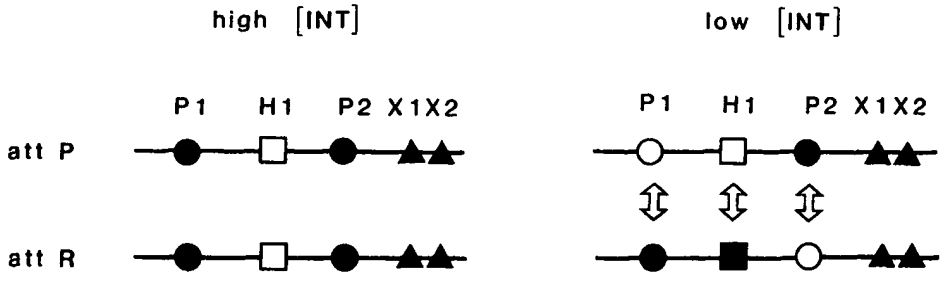


Figure 5 Coordinate occupancy of protein binding sites in the P arm of *att*R and *att*P as a result of cooperative and long-range interactions. Filled symbols indicate occupancy of the site by its cognate protein (P1 and P2 by Int, H1 by IHF, X1X2 by Xis). Double-headed arrows highlight the differences between *att*P and *att*R.

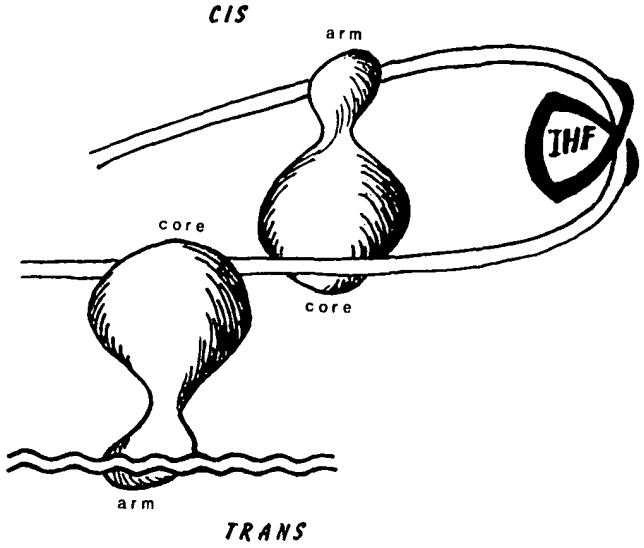


Figure 6 Basic design element of an intasome. IHF induces a sharp (≥140°) bend in DNA to bring the distant arm-type and core-type sites close enough to be bound by an Int protomer (*cis* interaction). The bivalent Int can also be involved in *(trans)* interactions between the arm-type and core-type sites on different partners. Xis and FIS are postulated to play a similar role to IHF.

SYNAPSIS AND STRAND EXCHANGE

Covalent Intermediates

The sites of strand exchange, i.e. the locus of cutting and resealing individual DNA strands, was narrowed down to two adjacent phosphodiester bonds, in an elegant isotope transfer experiment by M. Mizuuchi and K. Mizuuchi (171). Resolution of the remaining ambiguity came from sequence comparisons of many secondary *att* sites (3, 50) and from the demonstration by Craig & Nash (51) that Int makes a covalent link with the 3' phosphate at the site(s) of strand exchange (reviewed in Ref. 3).

In order to carry the analysis of transient recombination intermediates further, two families of suicide recombination substrates proved useful (see Figure 3) (13, 15). These *att* sites, which contain a strategically placed nick within the overlap region, initiate recombination normally and become deviant only after Int has acted on them. They are unable to complete or reverse the reaction. The particular intermediate trapped by such substrates depends upon where the nick is placed within the overlap region. One family of suicide substrates contains a medial nick in the overlap region (centered between the two sites of strand exchange) and leads to the formation of covalent Int-DNA

complexes in high yield. This work showed that Tyr-342 of Int forms a covalent link with DNA as an intermediate during strand cleavage and religation (15). The other family of suicide substrates is discussed below.

The Order of Strand Exchange

The existence of covalent Int-DNA intermediates does not distinguish between the two possible mechanisms of strand exchange: a concerted mechanism, in which both DNA strands in each *att* site are cleaved prior to strand exchange; and a sequential mechanism, in which one DNA strand in each *att* site is cleaved and exchanged followed by a second cycle of cleavage and exchange (see Figure 1).

From genetic crosses, it was found that Int-promoted recombination in vivo can generate the progeny predicted by a sequential strand exchange mechanism (172, 173). However, the low yield of these progeny was cited as evidence against their being on the primary recombination pathway (128). In vitro data that were interpreted to favor a concerted mechanism included the observation of low levels of double strand breakage by Int protein (128, 174). Support for a sequential strand exchange mechanism came from the specific resolution of synthetic *att* site Holliday junctions by purified Int protein to yield completed recombination products (12). The high efficiency and specificity of the resolution reaction was persuasive but not compelling.

Two different experimental devices finally led to the isolation of the elusive single-strand exchange recombination intermediates. One of these is the second family of suicide recombination substrates discussed above. In this family of suicide substrates, the preexisting *att* site nick is located at either the left (top strand) or right (bottom strand) Int cleavage site of the *att*B partner (see Figure 3). When the nick is in the bottom strand, single-strand exchange intermediates are produced under normal reaction conditions with almost the same efficiency as recombinant products from normal substrates. However, when the nick is in the top strand, no intermediates or recombinant products are formed (13). Thus, recombination proceeds via a sequential mechanism of strand exchange and the order is fixed such that the top strands must exchange first. Blocking the top strand exchange with a nick prevents the reaction from initiating. Using suicide recombination substrates for an *att*L × *att*R reaction, it was shown that excisive recombination has the same prescribed order of top strand exchange followed by bottom strand exchange (13, 14).

Similar conclusions have been drawn for integrative recombination (excision was not tested) from experiments in which substitution by phosphorothioate nucleotides into *att* site DNA was used to block Int cleavage (14, 65). Substitution in the top strand blocked strand exchange, while substitution in the bottom strand generated Holliday intermediates. In these experiments, however, very low levels of the recombination intermediate were seen, even

when using the Int variant known as Int-h (see above) to increase their yield. This difference between the analogue substitution and suicide recombination substrates is informative, especially when coupled with the low levels of Holliday structure obtained in normal reactions, or those blocked by a heterology in the overlap region (see below) (13, 14, 18). These differences have led to the suggestion that the first strand exchange during integrative recombination is reversible. Thus, in order to obtain high yields of recombination intermediate, it is not sufficient to block forward progression of the reaction, but it is also necessary to block its reversal (13). Understanding which of several possible mechanisms is responsible for the ability of the suicide substrate to trap Holliday junction intermediates is likely to provide useful insights into the normal reaction. Thus far, results obtained with a variety of different suicide recombination substrates indicate that Int protein, when provided with an appropriate 5' OH acceptor, will efficiently execute a single DNA strand transfer reaction (13) (S. E. Nunes-Düby, A. Landy, unpublished results).

The strict order of strand exchanges does not arise from asymmetry in the core region, but rather from the asymmetric arrangement of proteins bound to the P and P' arm sequences: (*a*) the core regions in the suicide recombination substrates were constructed to be virtually symmetric (13); (*b*) the same order of strand exchanges is obtained with *att* sites having different core region sequences (13, 14); (*c*) inversion of the overlap (13) or the full core region (14) does not change the order of strand exchange relative to the P and P' arms.

Holliday recombination intermediates have also been detected in the Cre and FLP systems (141, 145, 175). Thus, it is likely that this mechanism of sequential strand exchanges is common to members of the Int family.

DNA Homology

Although the specific sequence of the overlap region is not critical (within a wide latitude) for the order of strand exchange, or for the overall efficiency of recombination, it is required that the overlap region have the same sequence in both recombination partners (16, 17, 136). Two models have been proposed for the role of DNA : DNA homology in the overlap region. In the first, homology is required for synapsis of the recombining partners, and more specifically for the formation of a four-stranded DNA structure (7, 176, 177). From this view of recombination, considerable attention has been focused on whether the interstrand interactions are likely to be of the type proposed by McGavin (178) or by Wilson (179), and the likely direction in which strands rotate around the four-sided block during recombination (166, 177). In the second model, which is strongly favored by recent data, DNA : DNA homology is not required for synapsis but rather for some later step, such as branch

migration of a Holliday junction (16). Experiments with *att* sites that are heteroduplex within the overlap region argue against mechanisms involving a homology-dependent annealing of cohesive ends (180), however they do not distinguish between these two models.

A distinction between the two models is afforded by the experimental devices described above for obtaining single-strand exchange intermediates containing Holliday junctions. Whereas the formation of single-strand exchange intermediates is completely blocked by a nonhomology on the left side of the overlap region, it is not affected by a nonhomology on the right (13, 18). When the center of the overlap region is tested, single-strand exchange intermediate formation is blocked by a nonhomology at position 0 but not at position +1 (18). These results rule out that class of models, including that of four-stranded DNA, in which homology over all (or most) of the overlap region is required for synapsis. However, they do not exclude a requirement for a local homology of several base pairs in the immediate vicinity of a strand exchange site.

Whether or not strand exchange depends on some very short local homologies, a proposed requirement for a homology-dependent branch migration step is attractive (16). Artificial *att* site Holliday junctions have been made with sequence nonhomologies in the overlap region designed to restrict branch migration. For some of these constructions, resolution is strongly biased in favor of parental-type products, in contrast to wild-type Holliday structures that resolve to parental and recombinant-type products without pronounced bias. These results suggest that heterology prevents the Holliday junction from migrating from one Int cleavage site to the other and that resolution is executed by Int if the junction is at, or near, the Int cleavage site (B. DeMassy, L. Dorgai, R. A. Weisberg, personal communication). However, some data from these experiments suggest that Int may not be able to cleave when the junction is constrained very close to the cleavage site. The DNA branch migration model also gains further credibility from its ability to explain the results of recombination in which one partner is heteroduplex in the overlap region (17, 180).

Synapsis

Additional evidence that DNA homology is not required for synapsis comes from studies on both integrative and excisive recombination. Heteroduplex *att*B was incubated with a supercoiled homoduplex *att*P under recombination conditions; although the two *att* sites cannot recombine, because of nonhomologies between their overlap regions, double-strand cleavage of *att*B is promoted by the *att*P intasome. Thus, capture of *att*B by the *att*P intasome does not depend upon DNA homology between the sites (19). In addition, chemical footprinting assays establish that under recombination conditions,

*att*B cannot stably bind Int in competition with other DNAs. Taken together, these experiments suggest that *att*B may obtain its Int by binding to a preformed *att*P intasome in a synapsis that is governed primarily by protein-protein and protein-DNA interactions and is independent of DNA homology (19). Thus the asymmetry in structural complexity is indeed reflected in functional asymmetry. For the analysis of excisive recombination, half-*att* sites were constructed by cutting at a restriction site within the overlap region. These half-*att* sites are proficient for synapsis and strand transfer to an intact *att* site partner despite the lack of homology between their overlap regions (S. E. Nunes-Düby, A. Landy, unpublished results).

A particularly challenging aspect of synapsis is how two recombining *att* sites initially come together and are productively oriented with respect to one another. The fact that recombination can occur intermolecularly, intramolecularly, and between directly repeated or inverted sites suggests that *att* sites come together by random collision rather than by some form of tracking along the DNA. This general view is substantiated by topological analyses of the knots and catenanes generated by recombination: the topological complexity of recombinant products increases with both the density of supercoils and the distance between the *att* sites (130, 132, 164, 166, 181).

In contrast to the Int family, the Resolvase-Invertase family has more stringent orientation requirements and yields topologically invariant recombination products (33, 34, 39, 40). These features are consistent with a class of models in which the two recombination sites are interwrapped in a very specific manner, a requirement that is equivalent to a "topological filter" (182–185). From the topological experiments, it is clear that λ att sites do not have such an interwrapping requirement for productive synapsis. [Although an interesting exception may be the Xis-independent attL × attR recombination in low salt (84).] The usefulness of topological analyses of recombinant products lies mainly in its ability to set limits on acceptable models (40, 130, 166) when the number of biochemical unknowns is small. However, any acceptable mechanism for synapsis and strand exchange must be validated by its ability to explain the topological consequences of recombination.

Several features of the λ system are quite similar to those found in FLP and Cre, while others appear to be influenced by the unique arms of λ *att* sites. Both FLP (68, 186–188) and Cre (189–191) share with λ all those properties that suggest a random collision-type of synapsis. Cre, however, appears to exhibit an additional constraint in synapsis, suggesting some interwrapping of *lox* sites (189, 191, 192). However, it is unlikely that this aspect of synapsis reflects a fundamental difference in mechanism, since relatively small mutational changes in either Cre or *lox* can make the Cre system topologically similar to FLP and Int (192).

DIRECTIONALITY AND REGULATION

One of the unique and intriguing aspects of the site-specific recombination of λ (and related phages) is its extreme directionality. Although the pathway is commonly diagrammed as a reversible integration/excision reaction, and the normal products of one reaction are the normal substrates of the other, such notation belies the fact that there are in fact two distinct pathways. Integrative recombination normally depends upon supercoiling for assembly of the attP intasome (22, 169), while the attL and attR intasomes do not have this requirement (although they are stimulated by it) (133). Integrative recombination requires higher levels of Int than excisive recombination, both in vivo (193) and in vitro (when *att*R is supercoiled) (27, S. E. Nunes-Düby, A. Landy, unpublished results). Xis protein is required for excisive recombination under physiological conditions, while it is inhibitory for integrative recombination (80, 83, 194). The H1 site must be occupied by IHF for integrative recombination and it must be vacant for excision (27, 28, 88, 154). Of a total constellation of 13 protein binding sites, P1, H1, and P'3 are uniquely required for integrative recombination, and P2, X1, X2, F, and P'1 are uniquely required for, or involved in, excisive recombination (27–29, 87, 88, 153). Indeed, deletion of the P1 and H1 sites, which are required for *att*P function, makes an *att*R that can be more efficient than wild type (27). It is clear that the λ *att* sites have evolved to efficiently execute one reaction or the other in a mutually exclusive manner.

The regulation of phage integration and prophage excision consists of a sophisticated network of pathways involving multiple phage and host genes. A detailed description of how these pathways direct recombination, and some models for their regulation, have recently been presented (4). Several features of these models are summarized below.

In contrast to the analysis of gene expression, regulation at the level of the recombination reaction itself has been difficult to study in vivo. The in vitro data, and resulting models pointing to mechanism-based regulation, have to be tested by in vivo experiments specifically designed to detect such effects. For example, the large effect of FIS on in vivo excision, discussed above, was not detected until specifically looked for. The other mechanism-based regulation by a host-encoded protein involves IHF and the H1 binding site. The effect of IHF inhibition of excision is to greatly sensitize the reaction of Int concentration via the cooperative and competitive interactions described above (28, 88). Under one set of in vitro conditions, IHF inhibition at H1 sharpens the Int concentration dependence of excision by more than 10-fold (28).

The two host-encoded proteins involved in λ site-specific recombination

both exert regulatory effects on excision; FIS stimulates and IHF inhibits. It is significant that both regulatory responses are overridden by high levels of phage-encoded proteins: high Xis levels abolish FIS stimulation and high Int levels overcome IHF inhibition. These mechanisms endow the prophage excision reaction with a responsiveness to host physiology, viral physiology, and a range of environmental conditions (4, 28, 29). When there is a strong inducing stimulus, such as heavy UV irradiation, death of the host cell is imminent and it is advantageous for the prophage to excise and initiate a lytic cycle without delay or modulation. These are the conditions when levels of Int and Xis are expected to be high. When the inducing stimulus is weak and cell death is not certain, it is advantageous for viral excision and lytic development to be influenced by other factors, such as the likelihood of completing a lytic cycle. These are the conditions when Int and Xis levels may be low. Additionally, in the unfavorable conditions of stationary phase, FIS levels are low and excision is more difficult (requires higher levels of Xis).

The ability of FIS to stimulate excision at low levels of Xis may also be relevant in the phenomenon of spontaneous phage production, which occurs at a frequency of one in 10^3 to 10^5 cell generations. Some "spontaneous" phage production probably results from DNA damage and a strong induction in a small fraction of cells. Additionally, some cells may produce phage as a consequence of random fluctuations in repressor concentration, leading to a weak induction. It has been suggested (4) that the possible role of FIS in assisting this low-frequency pathway is not unlike the FIS stimulation of recombination of other low-frequency recombinations (112, 113, 115, 116). Both the *hin* inversion system of *Salmonella* flagellar antigens (195) and the *gin* inversion system of bacteriophage Mu tail fibers (196) also involve fluctuations in the synthesis of a poorly expressed recombinase. The stimulation of all three low-frequency reactions by FIS suggests that they will occur preferentially in exponentially growing cells.

COMMENTS AND SPECULATIONS

Cooperative Systems

In retrospect we can see why it was so difficult to isolate *att* site defective mutants (despite the existence of many potential targets for disruption). In highly cooperative complexes like intasomes, it is likely that small changes in specific protein-binding sites will be masked by the remaining functional sites. Indeed, even with the ability to make site-directed mutations, the phenotypes can be misleading. A two-base change in the consensus recognition sequence, which reduced IHF binding to the H1 site by 500-fold, yielded little change in phenotype. Only after the introduction of five base changes could it be determined that the H1 site is required for integrative recombina-

tion (28). Thus, in highly cooperative systems, considerable caution must be exercised in interpreting null phenotypes. Another example of confusing or misleading phenotypes was the observation that, under certain recombination conditions, deletion of the P2 Int binding site produced much less of a defect than a multisite mutant that destroyed its function. In fact, the interplay of competitive and cooperative binding in the P arm explains this surprising result, and is substantiated by biochemical analysis of the two mutants (87). A final example of considerations relevant to complex cooperative systems concerns the conditions under which apparently independent binding sites are actually used. The most prominent instances of this are the P arm binding site interactions (88), and the recent finding that some (or all) of the core-type sites do not obtain Int from solution (19), but through delivery by arm-type sites (L. Moitoso de Vargas, S. H. Kim, A. Landy, unpublished results).

We have also learned that optimization of an in vitro system can result in overlooking some of the regulatory and biochemical complexity that is important under nonoptimal conditions. Indeed, nonoptimal conditions are prevalent in nature. The effect of FIS protein could not have been detected in an in vitro system optimized for Xis concentration.

Dynamic Aspects

In considering how two *att* sites might synapse, it is useful to summarize the features of Int binding in the region of strand exchange. Both of the phosphodiester bonds that are cut during exchange of each strand, and all of the contacts for Int binding, are on the same face of the DNA helix, straddling the central major groove of the overlap region (50). The recent demonstration that synapsis does not require DNA homology across the entire 7 bp overlap region (13, 18, 19) rules out models of four-stranded DNA and eliminates the motivation for proposing extensive DNA-DNA interactions in the early steps of synapsis. This makes even more appealing the view that the Int proteins lie on the inside faces of a synaptic complex where they can interact with one another. One particularly attractive configuration is that of two helixes lying across one another (as opposed to a side-by-side alignment). This would serve to maximize the interaction between all of the bound Int proteins, as suggested by some preliminary data on the resolution of synthetic Holliday junctions and the properties of certain suicide recombination substrates (B. Franz, S. H. Kim, S. E. Nunes-Düby, A. Landy, unpublished results). It would also minimize the repulsive forces between the two DNA helixes. This view of the synaptic complex, with an "interior" cluster of Int molecules in the region of strand exchange, is similar to that suggested for γδ Resolvase on the basis of recent mutant and crystallographic data. These indicate a tetrameric complex of Resolvase monomers with the active-site serines on the outer surface (T. Steitz, N. Grindley, personal communication).

Structural Aspects

The sharp ($\cong$ 140°) bends induced at the loci of binding by IHF, Xis, and FIS, along with the bivalent DNA-binding capacity of Int, are probably the major determinants of intasome structure. Intrinsic DNA curvature, as a consequence of A tracts (124–126), is also likely to contribute to the overall structures. Within the intasome, these elements are likely to work in conjunction to form DNA loops that are tethered by an Int, binding simultaneously to arm-type and core-type sites. The delivery of Int to the low-affinity core-type sites, by the higher-affinity arm-type sites, would be facilitated by the protein-induced DNA bending. Binding at core-type sites could also be assisted through secondary (protein-protein) interactions. For example, one Int bound to both arm- and core-type sites could bring another Int to the core region, either on the same *att* site or on the synapsed partner. This view is also consistent with the cooperativity of Int binding in the core regions of all four *att* sites (50) (B. Franz, S. H. Kim, S. Nunes-Düby, L. Moitoso de Vargas, A. Landy, unpublished results).

This model of promoting binding to the core-type sites by the formation of DNA loops that deliver Int bound at arm-type sites predicts specific pairwise interactions between each core-type site and an arm-type site, or between two core-type sites. These interactions could involve sites within the same partner (*cis* interactions), and/or sites on different partners (*trans* interactions) (Figure 6). Similarly, the cooperative interactions between proteins bound at core-type sites could be *cis* and/or *trans*. Identifying these pairwise interactions is one of the challenges of understanding the intasome structures. A related challenge is characterization of the interactions between DNA and the site-specific DNA bending proteins IHF, Xis, FIS.

Regulatory Aspects

Bacteriophage λ has long served as a divining rod for uncovering host functions. Yet another example may be the host-encoded IHF and FIS proteins. It is tempting to speculate that the mechanism-based regulation of recombination incorporates IHF and FIS as integral elements of the reaction, because these two proteins reflect global signals of *E. coli* physiology. This suggestion is appealing because of the striking variation in FIS levels as a function of growth phase (29) and the wide range of reactions in which IHF is involved (111).

The role of FIS in stimulating excision has led to speculation about a new phage regulatory pathway that could maintain derepression even under weak inducing conditions (4, 29). After excision, new genes from the *b* region are expressed by N-dependent transcription through *att*P. One of these (*ben*) encodes a double-stranded DNA endonuclease (197, 198) that has been purified and found to cleave only supercoiled DNA (199). This endonuclease,

or one of the others encoded in the *b* region (22, 200, 201), could play a role in prolonging (inducing) an SOS response by cleaving the bacterial genome and generating new sites that activate RecA protein. The activated RecA would promote continuous cleavage of λ, thus preventing repression of the newly excised phage.

The model proposes a function for the enigmatic *b* region (202) and, more importantly, it suggests how FIS could enhance excision, even under conditions that might otherwise lead to repression of the excised phage. According to this model, spontaneous phage production does not reflect the lower limits of prophage repression, but rather is a programmed strategy in maintaining phage populations. In the absence of a pathway devoted to maintaining derepression (under weak inducing conditions), the dramatic effects of FIS on prophage excision would be primarily relevant to normal induction conditions.

CLOSING THE CIRCLE

Analysis of the λ recombination system has benefitted greatly from Robin Holliday's provocative insight about possible intermediates in the exchange (or rearrangement) of genetic information. In return, studies of the λ pathway have contributed significantly to understanding more about the formation, properties, and resolution of these genetic intermediates. The Holliday structures were conceived in the context of homologous, not site-specific, recombination—but such extrapolations are commonplace. One example is the possible relevance of intasome structures to the complex organizations involved in DNA replication, transcription, and other recombination systems. Understanding how the accessory proteins IHF, Xis, and FIS bend and shape DNA and how the bivalent Int protein tethers two distant DNA sites is likely to be of interest well beyond the field of site-specific recombination. Another example is the possible window into global regulatory devices that is afforded by the discovery of IHF and FIS proteins and their role in mechanism-based regulation of recombination.

The circle is always closing but it is never closed.

ACKNOWLEDGMENTS

I thank the many colleagues who provided information and manuscripts prior to publication. I am greatly indebted to the members of our research group, past and present, for their ideas, criticism and enthusiasm, especially to John Thompson, Simone Nunes-Düby, and Lina Moitoso de Vargas, who made significant contributions to this review. I also thank Joan Boyles for her dedicated and good-natured assistance in preparing the manuscript.

Literature Cited

1. Holliday, R. 1964. *Genet. Res.* 5:282–304
2. Nash, H. A. 1981. *Annu. Rev. Genet.* 15:143–67
3. Weisberg, R. A., Landy, A. 1983. In *Lambda II*, ed. F. W. Stahl, J. Roberts, R. A. Weisberg, pp. 211–50. New York: Cold Spring Harbor Lab.
4. Thompson, J. F., Landy, A. 1989. See Ref. 44, pp. 1–22
5. Campbell, A. M. 1962. *Adv. Genet.* 11:101–45
6. Kikuchi, Y., Nash, H. A. 1978. *J. Biol. Chem.* 253:7149–57
7. Kikuchi, Y., Nash, H. A. 1979. *Proc. Natl. Acad. Sci. USA* 76:3760–64
8. Nash, H. A., Robertson, C. A. 1981. *J. Biol. Chem.* 256:9246–53
9. Landy, A., Ross, W. 1977. *Science* 197:1147–60
10. Hsu, P.-L., Ross, W., Landy, A. 1980. *Nature* 285:85–91
11. Mizuuchi, M., Mizuuchi, K. 1980. *Proc. Natl. Acad. Sci. USA* 77:3220–24
12. Hsu, P.-L., Landy, A. 1984. *Nature* 311:721–26
13. Nunes-Düby, S. E., Matsumoto, L., Landy, A. 1987. *Cell* 50:779–88
14. Kitts, P. A., Nash, H. A. 1988. *J. Mol. Biol.* 204:95–108
15. Pargellis, C. A., Nunes-Düby, S. E., Moitoso de Vargas, L., Landy, A. 1988. *J. Biol. Chem.* 263:7678–85
16. Weisberg, R. A., Enquist, L. W., Foeller, C., Landy, A. 1983. *J. Mol. Biol.* 170:319–42
17. Bauer, C. E., Gardner, J. F., Gumport, R. I. 1985. *J. Mol. Biol.* 181:187–97
18. Kitts, P. A., Nash, H. A. 1987. *Nature* 329:346–48
19. Richet, E., Abcarian, P., Nash, H. A. 1988. *Cell* 52:9–17
20. Better, M., Lu, C., Williams, R. C., Echols, H. 1982. *Proc. Natl. Acad. Sci. USA* 79:5837–41
21. Better, M., Wickner, S., Auerbach, J., Echols, H. 1983. *Cell* 32:161–68
22. Richet, E., Abcarian, P., Nash, H. A. 1986. *Cell* 46:1011–21
23. Thompson, J. F., Landy, A. 1988. *Nucleic Acids Res.* 16:9687–705
24. Moitoso de Vargas, L., Pargellis, C. A., Hasan, N. M., Bushman, E. W., Landy, A. 1988. *Cell* 54:923–29
25. Argos, W., Landy, A., Abremski, K., Egan, J. B., Haggård-Ljungquist, E., et al. 1986. *EMBO J.* 5:433–40
26. Bushman, W., Yin, S., Thio, L. L., Landy, A. 1984. *Cell* 39:699–706
27. Bushman, W., Thompson, J. F., Vargas, L., Landy, A. 1985. *Science* 230:906–11
28. Thompson, J. F., Waechter-Brulla, D., Gumport, R. I., Gardner, J. F., Moitoso de Vargas, L., Landy, A. 1986. *J. Bacteriol.* 168:1343–51
29. Thompson, J. F., Moitoso de Vargas, L., Koch, C., Kahmann, R., Landy, A. 1987. *Cell* 50:901–8
30. Sadowski, P. 1986. *J. Bacteriol.* 165:341–47
31. Craig, N. L. 1988. *Annu. Rev. Genet.* 22:77–106
32. Hatfull, G. F., Salvo, J. J., Falvey, E. E., Rimphanitchayakit, V., Grindley, N. D. F. 1988. See Ref. 42, pp. 149–81
33. Hatfull, G. F., Grindley, N. D. F. 1988. See Ref. 43, pp. 357–96
34. Glasgow, A. C., Hughes, K. T., Simon, M. I. 1989. See. Ref. 44, pp. 637–59
35. Johnson, R. C., Simon, M. I. 1987. *Trends Genet.* 3:262–67
36. West, S. C. 1989. In *Nucleic Acids and Molecular Biology*, Vol. 3, ed. F. Eckstein, D. M. J. Lilley. Berlin: Springer Verlag. In press
37. West, S. C. 1988. *Trends Genet.* 4:8–13
38. Cox, M. M. 1989. See Ref. 44, pp. 661–70
39. Gellert, M., Nash, H. 1987. *Nature* 325:401–4
40. Benjamin, H. W., Cozzarelli, N. R. 1985. In *Proc. Robert A. Welch Found. Conf. Chem. Res. XXIX. Genetic Chemistry and the Molecular Basis of Heredity*, pp. 107–26. Houston: Robert A. Welch Found.
41. Wasserman, S. A., Cozzarelli, N. R. 1986. *Science* 232:951–60
42. Kingsman, A. J., Kingsman, S. M., Chater, K. F., eds. 1988. *Transposition*. Cambridge, England: Cambridge Univ. Press
43. Kucherlapati, R., Smith, G. R., eds. 1988. *Genetic Recombination*. Washington, DC: Am. Soc. Microbiol.
44. Berg, D. E., Howe, M. M., eds. 1989. *Mobile DNA*. Washington, DC: Am. Soc. Microbiol.
45. Hoess, R. H., Foeller, C., Bidwell, K., Landy, A. 1980. *Proc. Natl. Acad. Sci. USA* 77:2482–86
46. Davies, R. W. 1980. *Nucleic Acids Res.* 8:1765–82
47. Kotewicz, M., Grzesiuk, E., Courschesne, W., Fischer, R., Echols, H. 1980. *J. Biol. Chem.* 255:2433–39
48. Nash, H. A. 1981. *Enzymes* 14:471–80
49. Ross, W., Landy, A. 1982. *Proc. Natl. Acad. Sci. USA* 79:7724–28
50. Ross, W., Landy, A. 1983. *Cell* 33:261–72
51. Craig, N. L., Nash, H. A. 1983. *Cell* 35:795–803

52. Lange-Gustafson, B. J., Nash, H. A. 1984. *J. Biol. Chem.* 259:12724–32
53. Belfort, M. 1980. *Gene* 11:149–55
54. Schindler, D., Echols, H. 1981. *Proc. Natl. Acad. Sci. USA* 78:4475–79
55. Guarneros, G., Montanez, C., Hernandez, T., Court, D. 1982. *Proc. Natl. Acad. Sci. USA* 79:238–42
56. Schmeissner, U., McKenney, K., Rosenberg, M., Court, D. 1984. *J. Mol. Biol.* 176:39–53
57. Montanez, C., Bueno, J., Schmeissner, U., Court, D. L., Guarneros, G. 1986. *J. Mol. Biol.* 191:29–37
58. Shimada, K., Campbell, A. 1974. *Proc. Natl. Acad. Sci. USA* 71:237–41
59. Leong, J., Nunes-Düby, S., Oser, A. B., Lesser, C., Youderian, P., et al. 1986. *J. Mol. Biol.* 189:603–16
60. Pierson, L. S. III, Kahn, M. L. 1987. *J. Mol. Biol.* 196:487–96
61. Yu, A., Bertani, L. E., Haggård-Ljungquist, E. 1989. *Gene*. In press
62. Ljungquist, E., Bertani, L. E. 1983. *Mol. Gen. Genet.* 192:87–94
63. Bear, S. E., Clemens, J. B., Enquist, L. W., Zagursky, R. J. 1987. *J. Bacteriol.* 169:5880–83
64. Miller, H. I., Mozola, M. A., Friedman, D. I. 1980. *Cell* 20:721–29
65. Kitts, P. A., Nash, H. A. 1988. *Nucleic Acids Res.* 16:6839–56
66. Enquist, L. W., Weisberg, R. A. 1984. *Mol. Gen. Genet.* 195:62–69
67. Dayhoff, M. O., Schwartz, R. M., Orcott, B. L. 1978. In *Atlas of Protein Sequence and Structure*, ed. M. O. Dayhoff, pp. 345–52. Washington, DC: Natl. Biomed. Res. Found.
68. Gronostajski, R. M., Sadowski, P. D. 1985. *Mol. Cell. Biol.* 5:3274–79
69. Prasad, P. V., Young, L.-J., Jayaram, M. 1987. *Proc. Natl. Acad. Sci. USA* 84:2189–93
70. Wierzbicki, A., Kendall, M., Abremski, K., Hoess, R. 1987. *J. Mol. Biol.* 195:785–94
71. Reed, R. R., Moser, C. D. 1984. *Cold Spring Harbor Symp. Quant. Biol.* 49:245–49
72. Hatfull, G. F., Grindley, N. D. F. 1986. *Proc. Natl. Acad. Sci. USA* 83:5429–33
73. Klippel, A., Mertens, G., Patschinsky, T., Kahmann, R. 1988. *EMBO J.* 7: 1229–37
74. Ross, W., Landy, A., Kikuchi, Y., Nash, H. 1979. *Cell* 18:297–307
75. Nussinov, R., Weisberg, R. A. 1986. *J. Biomol. Struct. Dyn.* 3:1133–44
76. Enquist, L. W., Weisberg, R. A. 1977. *J. Mol. Biol.* 111:97–120
77. Gingery, R., Echols, H. 1967. *Proc. Natl. Acad. Sci. USA* 58:1507–14
78. Gottesman, M. E., Yarmolinsky, M. 1968. *J. Mol. Biol.* 31:487–505
79. Zissler, J. 1967. *Virology* 31:189
80. Abremski, K., Gottesman, S. 1982. *J. Biol. Chem.* 257:9658–62
81. Weisberg, R. A., Gottesman, M. E. 1971. In *Bacteriophage Lambda*, ed. A. D. Hershey, pp. 489–500. New York: Cold Spring Harbor Lab. Press
82. Gottesman, S., Gottesman, M., Shaw, J. E., Pearson, M. L. 1981. *Cell* 24:225–33
83. Abremski, K., Gottesman, S. 1981. *J. Mol. Biol.* 153:67–78
84. Craig, N. L., Nash, H. A. 1983. In *Mechanism of DNA Replication and Recombination*, ed. N. R. Cozzarelli, pp. 617–36. New York: Liss
85. Thompson, J. F., Snyder, U. K., Landy, A. 1988. *Proc. Natl. Acad. Sci. USA* 85:6323–27
86. Yin, S., Bushman, W., Landy, A. 1985. *Proc. Natl. Acad. Sci. USA* 82:1040–44
87. Thompson, J. F., Moitoso de Vargas, L., Nunes-Düby, S. E., Pargellis, C., Skinner, S. E., Landy, A. 1987. In *Mechanisms of DNA Replication and Recombination*, ed. T. Kelly, R. McMacken, pp. 735–44. New York: Liss
88. Thompson, J. F., Moitoso de Vargas, L., Skinner, S. E., Landy, A. 1987. *J. Mol. Biol.* 195:481–93
89. Miller, H. I., Friedman, D. I. 1980. *Cell* 20:711–19
90. Kikuchi, A., Flamm, E., Weisberg, R. A. 1985. *J. Mol. Biol.* 183:129–40
91. Miller, H. I., Kikuchi, A., Nash, H. A., Weisberg, R. A., Friedman, D. I. 1979. *Cold Spring Harbor Symp. Quant. Biol.* 43:1121–26
92. Miller, H. I., Nash, H. A. 1981. *Nature* 290:523–26
93. Miller, H. I. 1984. *Cold Spring Harbor Symp. Quant. Biol.* 49:691–98
94. Flamm, E. L., Weisberg, R. A. 1985. *J. Mol. Biol.* 183:117–28
95. Mechulam, Y., Fayat, G., Blanquet, S. 1985. *J. Bacteriol.* 163:787–91
96. Drlica, K., Rouviere-Yaniv, J. 1987. *Microbiol. Rev.* 51:301–19
97. Craig, N. L., Nash, H. A. 1984. *Cell* 39:707–16
98. Leong, J. M., Nunes-Düby, S., Lesser, C. F., Youderian, P., Susskind, M. M., Landy, A. 1985. *J. Biol. Chem.* 260:4468–77
99. Gamas, P., Chandler, M. G., Prentki, P., Galas, D. J. 1987. *J. Mol. Biol.* 195:261–72
100. Stenzel, T. T., Patel, P., Bastia, D. 1987. *Cell* 49:709–17
101. Prentki, P., Chandler, M., Galas, D. J. 1987. *EMBO J.* 6:2479–87

102. Robertson, C. A., Nash, H. A. 1988., *J. Biol. Chem.* 263:3554–57
103. Kosturko, L. D., Daub, E., Murialdo, H. 1989. *Nucleic Acids Res.* 17:17–34
104. Tanaka, I., Appelt, K., Dijk, J., White, S. W., Wilson, K. S. 1984. *Nature* 310:376–81
105. Friedrich, M. J., DeVeaux, L. C., Kadner, R. J. 1986. *J. Bacteriol.* 167:928–34
106. Nash, H. A., Robertson, C. A., Flamm, E., Weisberg, R. A., Miller, H. I. 1987. *J. Bacteriol.* 169:4124–27
107. Miller, H. I., Friedman, D. I. 1977. In *DNA Insertion Elements*, ed. A. I. Bukhari, J. A. Shapiro, S. L. Adhya, pp. 349–56. New York: Cold Spring Harbor Lab.
108. Morisato, D., Kleckner, N. 1987. *Cell* 51:101–11
109. Way, J. C., Kleckner, N. 1986. *Proc. Natl. Acad. Sci. USA* 81:3452–56
110. Surette, M. G., Chaconas, G. 1989. *J. Biol. Chem.* 264:3028–34
111. Friedman, D. I. 1988. *Cell* 55:545–54
112. Johnson, R. C., Simon, M. I. 1985. *Cell* 41:781–91
113. Kahmann, R., Rudt, F., Koch, C., Mertens, G. 1985. *Cell* 41:771–80
114. Huber, H. E., Iida, S., Arber, W., Bickle, T. A. 1985. *Proc. Natl. Acad. Sci. USA* 82:3776–80
115. Johnson, R. C., Bruist, M. F., Simon, M. I. 1986. *Cell* 46:531–39
116. Koch, C., Kahmann, R. 1986. *J. Biol. Chem.* 261:15673–78
117. Johnson, R. C., Ball, C. A., Pfeffer, D., Simon, M. I. 1988. *Proc. Natl. Acad. Sci. USA* 85:3484–88
118. Koch, C., Vandekerckhove, J., Kahmann, R. 1988. *Proc. Natl. Acad. Sci. USA* 85:4237–41
119. Pabo, C., Sauer, R. 1984. *Annu. Rev. Biochem.* 53:293–321
120. Plasterk, R. H. A., van de Putte, P. 1984. *Biochim. Biophys. Acta* 782: 111–19
121. Johnson, R. C., Glasgow, A. C., Simon, M. I. 1987. *Nature* 329:462–65
122. Bruist, M. F., Glasgow, A. C., Johnson, R. C., Simon, M. I. 1987. *Genes Dev.* 1:762–72
123. Mizuuchi, M., Mizuuchi, K. 1985. *Nucleic Acids Res.* 13:1193–208
124. Marini, J. C., Levene, S. D., Crothers, D. M., Englund, P. T. 1982. *Proc. Natl. Acad. Sci. USA* 79:7664–68
125. Wu, H.-M., Crothers, D. M. 1984. *Nature* 308:509–13
126. Diekmann, S., Wang, J. C. 1985. *J. Mol. Biol.* 186:1–11
127. Thompson, J. F., Mark, H. F., Franz, B., Landy, A. 1988. In *DNA Bending and Curvature*, ed. W. K. Olson, M. H. Sarma, R. H. Sarma, M. Sundaralingam, pp. 119–28. Guilderland, NY: Adenine
128. Kitts, P., Richet, E., Nash, H. A. 1984. *Cold Spring Harbor Symp. Quant. Biol.* 49:735–44
129. Laundon, C. H., Griffith, J. D. 1988. *Cell* 52:545–49
130. Spengler, S. J., Stasiak, A., Cozzarelli, N. R. 1985. *Cell* 42:325–34
131. Mizuuchi, K., Gellert, M., Nash, H. 1978. *J. Mol. Biol.* 121:375–92
132. Pollock, T. J., Nash, H. A. 1983. *J. Mol. Biol.* 170:1–18
133. Abremski, K., Gottesman, S. 1979. *J. Mol. Biol.* 131:637–49
134. Shulman, M., Gottesman, M. 1973. *J. Mol. Biol.* 81:461–82
135. Ross, W., Landy, A. 1982. *J. Mol. Biol.* 156:505–29
136. DeMassy, B., Studier, F. W., Dorgai, L., Appelbaum, E., Weisberg, R. A. 1984. *Cold Spring Harbor Symp. Quant. Biol.* 49:715–26
137. Hoess, R. H., Abremski, K. 1985. *J. Mol. Biol.* 181:351–62
138. Hoess, R., Abremski, K., Sternberg, N. 1984. *Cold Spring Harbor Symp. Quant. Biol.* 49:761–68
139. Hoess, R. H., Wierzbicki, A., Abremski, K. 1986. *Nucleic Acids Res.* 14: 2287–300
140. Abremski, K., Wierzbicki, A., Frommer, B., Hoess, R. H. 1986. *J. Biol. Chem.* 261:391–96
141. Hoess, R., Wierzbicki, A., Abremski, K. 1987. *Proc. Natl. Acad. Sci. USA* 84:6840–44
142. Andrews, B. J., Beatty, L. G., Sadowski, P. D. 1987. *J. Mol. Biol.* 193:345–58
143. Senecoff, J. F., Cox, M. M. 1986. *J. Biol. Chem.* 261:7380–86
144. Umlauf, S. W., Cox, M. M. 1988. *EMBO J.* 7:1845–52
145. Meyer-Leon, L., Huang, L.-C., Umlauf, S. W., Cox, M. M., Inman, R. B. 1988. *Mol. Cell. Biol.* 8:3784–96
146. Andrews, B. J., McLeod, M., Broach, J. R., Sadowski, P. D. 1986. *Mol. Cell. Biol.* 6:2482–89
147. Iida, S., Hiestand-Nauer, R. 1987. *Mol. Gen. Genet.* 208:464–68
148. Reed, R. R., Grindley, N. D. F. 1981. *Cell* 25:721–28
149. Shimada, K., Weisberg, R. A., Gottesman, M. E. 1972. *J. Mol. Biol.* 63:483–503
150. Shimada, K., Weisberg, R. A., Gottesman, M. E. 1972. *J. Mol. Biol.* 80:297–314
151. Shimada, K., Weisberg, R. A., Gottesman, M. E. 1975. *J. Mol. Biol.* 93:415–29

152. Winoto, A., Chung, S., Abraham, J., Echols, H. 1986. *J. Mol. Biol.* 192:677–80
153. Bauer, C. E., Hesse, S. D., Gumport, R. I., Gardner, J. F. 1986. *J. Mol. Biol.* 192:513–27
154. Gardner, J. F., Nash, H. A. 1986. *J. Mol. Biol.* 191:181–89
155. Krause, H. M., Higgins, N. P. 1986. *J. Biol. Chem.* 261:3744–52
156. Filutowicz, M., Appelt, K. 1988. *Nucleic Acids Res.* 16:3829–32
157. Gartenberg, M. R., Crothers, D. M. 1988. *Nature* 333:824–29
158. Davies, R. W., Schreier, P. H., Kotewicz, M. L., Echols, H. 1979. *Nucleic Acids Res.* 7:2255–73
159. Rhodes, D., Klug, A. 1980. *Nature* 286:573–78
160. Klug, A., Lutter, L. C. 1981. *Nucleic Acids Res.* 9:4267–83
161. Liu, L. F., Wang, J. 1978. *Cell* 15:979–84
162. Drew, H. R., Travers, A. A. 1985. *J. Mol. Biol.* 186:773–90
163. Hamilton, D., Yuan, R., Kikuchi, Y. 1981. *J. Mol. Biol.* 152:163–69
164. Mizuuchi, K., Gellert, M., Weisberg, R. A., Nash, H. A. 1980. *J. Mol. Biol.* 141:485–95
165. Mizuuchi, K., Fisher, L. M., O'Dea, M. H., Gellert, M. 1980. *Proc. Natl. Acad. Sci. USA* 72:1847–51
166. Griffith, J. D., Nash, H. A. 1985. *Proc. Natl. Acad. Sci. USA* 82:3124–28
167. Germond, J. E., Hirt, B., Oudet, P., Gross-Bellard, M., Chambon, P. 1975. *Proc. Natl. Acad. Sci. USA* 72:1843–47
168. Richmond, T. J., Finch, J. T., Klug, A. 1983. *Cold Spring Harbor Symp. Quant. Biol.* 47:493–501
169. Mizuuchi, K., Mizuuchi, M. 1979. *Cold Spring Harbor Symp. Quant. Biol.* 43:1111–14
170. Pollock, T. J., Abremski, K. 1979. *J. Mol. Biol.* 131:651–54
171. Mizuuchi, K., Weisberg, R., Enquist, L., Mizuuchi, M., Buraczynska, M., et al. 1981. *Cold Spring Harbor Symp. Quant. Biol.* 45:429–37
172. Echols, H., Green, L. 1979. *Genetics* 93:297–307
173. Enquist, L. W., Nash, H., Weisberg, R. A. 1979. *Proc. Natl. Acad. Sci. USA* 76:1363–67
174. Bauer, C. E., Hesse, S. D., Gardner, J. F., Gumport, R. I. 1984. *Cold Spring Harbor Symp. Quant. Biol.* 49:699–705
175. Jayaram, M., Crain, K. L., Parsons, R. L., Harshey, R. M. 1988. *Proc. Natl. Acad. Sci. USA* 85:7902–6
176. Nash, H. A., Mizuuchi, K., Enquist, L. W., Weisberg, R. A. 1981. *Cold Spring Harbor Symp. Quant. Biol.* 45:417–27
177. Nash, H. A., Pollock, T. J. 1983. *J. Mol. Biol.* 170:19–38
178. McGavin, S. 1971. *J. Mol. Biol.* 55:293–98
179. Wilson, J. H. 1979. *Proc. Natl. Acad. Sci. USA* 76:3641–45
180. Nash, H. A., Bauer, C. E., Gardner, J. F. 1987. *Proc. Natl. Acad. Sci. USA* 84:4049–53
181. Bliska, J. B., Cozzarelli, N. R. 1987. *J. Mol. Biol.* 197:205–18
182. Boocock, M. R., Brown, J. L., Sherratt, D. J. 1985. *Biochem. Soc. Trans.* 14: 214–16
183. Craigie, R., Mizuuchi, K. 1986. *Cell* 45:795–800
184. Wasserman, S. A., Cozzarelli, N. R. 1985. *Proc. Natl. Acad. Sci. USA* 82:1079–83
185. Wasserman, S. A., Dungan, J. M., Cozzarelli, N. R. 1985. *Science* 229:171–74
186. Meyer-Leon, L., Gates, C. A., Attwood, J. M., Wood, E. A., Cox, M. M. 1987. *Nucleic Acids Res.* 15:6469–88
187. Vetter, D., Andrews, B. J., Roberts-Beatty, L., Sadowski, P. D. 1983. *Proc. Natl. Acad. Sci. USA* 80:7284–88
188. Beatty, L. G., Babineau-Clary, P., Hogrefe, C., Sadowski, P. D. 1986. *J. Mol. Biol.* 188:529–44
189. Abremski, K., Hoess, R., Sternberg, N. 1983. *Cell* 32:1301–11
190. Abremski, K., Frommer, B., Hoess, R. H. 1986. *J. Mol. Biol.* 192:17–26
191. Abremski, K., Hoess, R. 1985. *J. Mol. Biol.* 184:211–20
192. Abremski, K., Frommer, B., Wierzbicki, A., Hoess, R. H. 1988. *J. Mol. Biol.* 201:1–8
193. Enquist, L. W., Kikuchi, A., Weisberg, R. A. 1979. *Cold Spring Harbor Symp. Quant. Biol.* 43:1115–20
194. Nash, H. A. 1975. *Proc. Natl. Acad. Sci. USA* 72:1072–76
195. van de Putte, P., Cramer, S., Giphart-Gassler, M. 1980. *Nature* 286:218–22
196. Zieg, J., Hillman, M., Simon, M. 1978. *Cell* 15:237–44
197. Benchimol, S., Lucko, H., Becker, A. 1982. *J. Biol. Chem.* 257:5201–10
198. Sumner-Smith, M., Benchimol, S., Murialdo, H., Becker, A. 1982. *J. Mol. Biol.* 160:1–22
199. Benchimol, S., Lucko, H., Becker, A. 1982. *J. Biol. Chem.* 257:5211–19
200. Becker, A. 1970. *Biochem. Biophys. Res. Commun.* 41:63–70
201. Chowdbury, M. R., Dunbar, S., Becker, A. 1972. *Virology* 49:314–18
202. Kellenberger, G., Zichichi, M. L., Weigle, J. 1961. *J. Mol. Biol.* 3:399–408

A


TnpI(1)
core 1.2 5′ TAATACAACACAAT ATTA ATTGTGTTGTATTA
3′ ATTATGTTGTGTTA TAAT TAACACAACATAAT
TnpI(2)
core 1.1 5′ TAATACAACACAAT ATAT ATTGTGTTGTATTA
3′ ATTATGTTGTGTTA TATA TAACACAACATAAT
core 2.2 5′ TAATACAACACAAT TATA ATTGTGTTGTATTA
3′ ATTATGTTGTGTTA ATAT TAACACAACATAAT

B

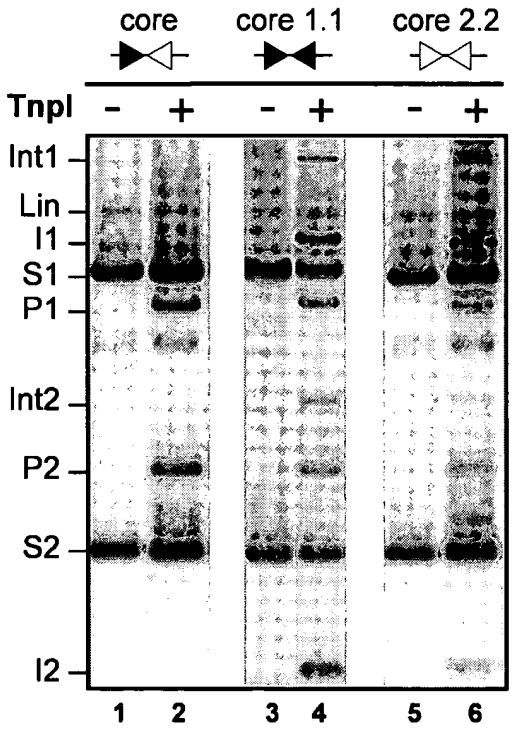

core
core 1.1
core 2.2
TnpI - + - + - +
Int1
Lin
I1
S1
P1
Int2
P2
S2
I2
1 2 3 4 5 6

Fig.1